

KW	termination sequence.	PR	28-JUN-1999;	99US-0140823.
XX		PR	29-JUN-1999;	99US-0140991.
OS	Arabidopsis thaliana.	PR	30-JUN-1999;	99US-0141287.
XX		PR	01-JUL-1999;	99US-0141842.
FN	EP1033405-A2.	PR	01-JUL-1999;	99US-0142154.
XX		PR	02-JUL-1999;	99US-0142055.
PD	06-SEP-2000.	PR	06-JUL-1999;	99US-0142390.
XX		PR	08-JUL-1999;	99US-0142803.
PF	25-FEB-2000; 2000EP-0301439.	PR	09-JUL-1999;	99US-0142920.
XX		PR	12-JUL-1999;	99US-0142977.
XX		PR	13-JUL-1999;	99US-0143542.
PR	25-FEB-1999;	PR	14-JUL-1999;	99US-0143624.
PR	05-MAR-1999;	PR	15-JUL-1999;	99US-0144005.
PR	09-MAR-1999;	PR	16-JUL-1999;	99US-0144085.
PR	23-MAR-1999;	PR	16-JUL-1999;	99US-0144086.
PR	25-MAR-1999;	PR	16-JUL-1999;	99US-0144086.
PR	29-MAR-1999;	PR	19-JUL-1999;	99US-0144325.
PR	01-APR-1999;	PR	19-JUL-1999;	99US-0144331.
PR	06-APR-1999;	PR	19-JUL-1999;	99US-0144332.
PR	08-APR-1999;	PR	19-JUL-1999;	99US-0144333.
PR	16-APR-1999;	PR	19-JUL-1999;	99US-0144334.
PR	19-APR-1999;	PR	19-JUL-1999;	99US-0144335.
PR	19-APR-1999;	PR	20-JUL-1999;	99US-0144352.
PR	21-APR-1999;	PR	20-JUL-1999;	99US-0144632.
PR	23-APR-1999;	PR	20-JUL-1999;	99US-0144884.
PR	28-APR-1999;	PR	21-JUL-1999;	99US-0144884.
PR	30-APR-1999;	PR	21-JUL-1999;	99US-0145086.
PR	30-APR-1999;	PR	21-JUL-1999;	99US-0145088.
PR	04-MAY-1999;	PR	22-JUL-1999;	99US-0145088.
PR	05-MAY-1999;	PR	22-JUL-1999;	99US-0145087.
PR	06-MAY-1999;	PR	22-JUL-1999;	99US-0145089.
PR	07-MAY-1999;	PR	23-JUL-1999;	99US-0145192.
PR	11-MAY-1999;	PR	23-JUL-1999;	99US-0145145.
PR	14-MAY-1999;	PR	23-JUL-1999;	99US-0145218.
PR	14-MAY-1999;	PR	26-JUL-1999;	99US-0145224.
PR	14-MAY-1999;	PR	26-JUL-1999;	99US-0145276.
PR	14-MAY-1999;	PR	27-JUL-1999;	99US-0145913.
PR	18-MAY-1999;	PR	27-JUL-1999;	99US-0145918.
PR	19-MAY-1999;	PR	27-JUL-1999;	99US-0145919.
PR	20-MAY-1999;	PR	28-JUL-1999;	99US-0145951.
PR	21-MAY-1999;	PR	28-JUL-1999;	99US-0146386.
PR	21-MAY-1999;	PR	02-AUG-1999;	99US-0146388.
PR	24-MAY-1999;	PR	02-AUG-1999;	99US-0146389.
PR	25-MAY-1999;	PR	03-AUG-1999;	99US-0147038.
PR	27-MAY-1999;	PR	04-AUG-1999;	99US-0147204.
PR	28-MAY-1999;	PR	04-AUG-1999;	99US-0147302.
PR	01-JUN-1999;	PR	05-AUG-1999;	99US-0147192.
PR	03-JUN-1999;	PR	05-AUG-1999;	99US-0147260.
PR	04-JUN-1999;	PR	06-AUG-1999;	99US-0147303.
PR	07-JUN-1999;	PR	06-AUG-1999;	99US-0147416.
PR	08-JUN-1999;	PR	09-AUG-1999;	99US-0147493.
PR	10-JUN-1999;	PR	09-AUG-1999;	99US-0147935.
PR	14-JUN-1999;	PR	10-AUG-1999;	99US-0148171.
PR	16-JUN-1999;	PR	11-AUG-1999;	99US-0148319.
PR	16-JUN-1999;	PR	12-AUG-1999;	99US-0148341.
PR	17-JUN-1999;	PR	13-AUG-1999;	99US-0148565.
PR	17-JUN-1999;	PR	13-AUG-1999;	99US-0148684.
PR	18-JUN-1999;	PR	16-AUG-1999;	99US-0149368.
PR	18-JUN-1999;	PR	17-AUG-1999;	99US-0149175.
PR	18-JUN-1999;	PR	18-AUG-1999;	99US-0149426.
PR	18-JUN-1999;	PR	20-AUG-1999;	99US-0149722.
PR	18-JUN-1999;	PR	20-AUG-1999;	99US-0149723.
PR	18-JUN-1999;	PR	20-AUG-1999;	99US-0149929.
PR	18-JUN-1999;	PR	23-AUG-1999;	99US-0149902.
PR	18-JUN-1999;	PR	23-AUG-1999;	99US-0149930.
PR	18-JUN-1999;	PR	25-AUG-1999;	99US-0150566.
PR	18-JUN-1999;	PR	26-AUG-1999;	99US-0150884.
PR	18-JUN-1999;	PR	27-AUG-1999;	99US-0151065.
PR	18-JUN-1999;	PR	27-AUG-1999;	99US-0151066.
PR	21-JUN-1999;	PR	27-AUG-1999;	99US-0151080.
PR	22-JUN-1999;	PR	30-AUG-1999;	99US-0151303.
PR	23-JUN-1999;	PR	31-AUG-1999;	99US-0151438.
PR	23-JUN-1999;	PR	01-SEP-1999;	99US-0151930.
PR	24-JUN-1999;	PR	07-SEP-1999;	99US-0152363.

PR	10-SEP-1999;	99US-0153070.
PR	13-SEP-1999;	99US-0153758.
PR	15-SEP-1999;	99US-0154018.
PR	16-SEP-1999;	99US-0154039.
PR	20-SEP-1999;	99US-0154779.
PR	22-SEP-1999;	99US-0155139.
PR	23-SEP-1999;	99US-0155486.
PR	24-SEP-1999;	99US-0155659.
PR	28-SEP-1999;	99US-0156458.
PR	29-SEP-1999;	99US-0156596.
PR	04-OCT-1999;	99US-0157117.
PR	05-OCT-1999;	99US-0157753.
PR	06-OCT-1999;	99US-0157865.
PR	07-OCT-1999;	99US-0158029.
PR	08-OCT-1999;	99US-0158232.
PR	12-OCT-1999;	99US-0158369.
PR	13-OCT-1999;	99US-0159293.
PR	13-OCT-1999;	99US-0159294.
PR	13-OCT-1999;	99US-0159295.
PR	14-OCT-1999;	99US-0159329.
PR	14-OCT-1999;	99US-0159330.
PR	14-OCT-1999;	99US-0159331.
PR	14-OCT-1999;	99US-0159637.
PR	14-OCT-1999;	99US-0159638.
PR	18-OCT-1999;	99US-0159584.
PR	21-OCT-1999;	99US-0160741.
PR	21-OCT-1999;	99US-0160767.
PR	21-OCT-1999;	99US-0160768.
PR	21-OCT-1999;	99US-0160770.
PR	21-OCT-1999;	99US-0160814.
PR	21-OCT-1999;	99US-0160815.
PR	22-OCT-1999;	99US-0160980.
PR	22-OCT-1999;	99US-0160981.
PR	22-OCT-1999;	99US-0160989.
PR	25-OCT-1999;	99US-0161404.
PR	25-OCT-1999;	99US-0161405.
PR	25-OCT-1999;	99US-0161406.
PR	26-OCT-1999;	99US-0161359.
PR	26-OCT-1999;	99US-0161360.
PR	26-OCT-1999;	99US-0161361.
PR	28-OCT-1999;	99US-0161920.
PR	28-OCT-1999;	99US-0161992.
PR	28-OCT-1999;	99US-0161993.
PR	29-OCT-1999;	99US-0161442.
Query Match 10.5%; Score 327; DB 21; Length 456;		
Best Local Similarity 25.6%; Pred. No. 1.6e-20;		
Matches 127; Conservative 84; Mismatches 198; Indels 88; Gaps 20;		
QY	85	STENATDSKLENGVLAARNNNPHQLSP-----PRN-----NYSVETTVTSQP- 131
Db	3	TINSALD-----KYNN--APVSKPSVGRWFPQNAVEDSSNVFAADSIDDNPD 48
QY	132	-----SDEIVP-----ETSNKPEGGRMCELMCSKGLRKTIYANISDQAKLMBEIPSAKLA 182
Db	49	GIVQDVQISPVNMGNEVKAYPQLLKLGDMDSGLHKFVSDNRKNLASLKEEIPMAFRAA 108
QY	183	KPAKVLDCIKFY-----LQGRATKESPMGSAQVSVLLILE--SFLIMP-DRGKG 233
Db	109	ANPASVLDSLEGFVPMEAPTADGX-----KDALNGMERTCIMLMCELSILLGLDNCNL 164
QY	234	KVKIESWIKDEAETAAVAKRELMT-EGGLAAAEKMDARGILLVACFGVPSNFRSTDLL 292
Db	165	AVVLSONVHKRATIAEGWNPLIESLDMDACNGNSLEAHAFIQLLATPAIIVADFKEDELL 224
QY	293	DLIRMSGNEIAGALKRSQFIIPWMSGVIVESSIKRGMHIEALEMWYIFGMEDKFSALVL 352
Db	225	KLIPMVSRRQAEALCRSLGLAEKMPGVIEVLVNSGKQIDAVNLAFAPELTEQSPSVSL 284
QY	353	TSEFLKMSKESFRKRAKQASPL--AFKBAATKQALVLSVNMCMETHKLDPAKELPGQWI 410
Db	285	KSVLIEARRSSQGRPNASPAVQASDENEREELGLKTVIKCIEHSLE--EQYP---- 338

QY	411	KEQIVSLEKOTLQDKEKMEKARSLSLMBEALAKRMVNOQIKRPRLSPEMPPVTSSS- 469
Db	339	---VEPLHKLRIQLEKAKADKKRATEPMK-----PQKRPGRGA---QPRVTDNN 382
QY	470	-----YSPIYDRDRFPSPQSDDDQDEISALYSSYLGPSTSPPHSRSRSPYMWPLPHG 521
Db	383	NINNNKTCYGRVIPER-YPCVYVDNRPFSLGPIMAAQPPPPPPQTYTNP---AP-ARG 437
QY	522	GLGRSVYAYEHLAPNSY 538
Db	438	NFYANCYQYQAPPPPPY 454
RESULT 8		
AAG11239		
ID	AAG11239 standard; Protein; 532 AA.	
XX	AAG11239;	
AC	AAG11239;	
XX	17-OCT-2000 (first entry)	
DT	17-OCT-2000 (first entry)	
XX	Arabidopsis thaliana protein fragment SEQ ID NO: 9882.	
DE	Arabidopsis thaliana	
XX	Protein identification; signal transduction pathway; metabolic pathway;	
KW	hybridisation assay; genetic mapping; gene expression control; promoter;	
KW	termination sequence.	
XX	Arabidopsis thaliana.	
OS	Arabidopsis thaliana.	
PN	EP1033405-A2.	
XX	06-SEP-2000.	
PD	06-SEP-2000.	
XX	25-FEB-2000; 2000EP-0301439.	
PF	25-FEB-2000; 2000EP-0301439.	
XX	25-FEB-1999; 99US-0121825.	
PR	05-MAR-1999; 99US-0123180.	
PR	09-MAR-1999; 99US-0123548.	
PR	23-MAR-1999; 99US-0125788.	
PR	25-MAR-1999; 99US-0126264.	
PR	29-MAR-1999; 99US-0126785.	
PR	01-APR-1999; 99US-0127462.	
PR	06-APR-1999; 99US-0128234.	
PR	08-APR-1999; 99US-0128714.	
PR	16-APR-1999; 99US-0129845.	
PR	19-APR-1999; 99US-0130077.	
PR	21-APR-1999; 99US-0130449.	
PR	23-APR-1999; 99US-0130510.	
PR	23-APR-1999; 99US-0130891.	
PR	28-APR-1999; 99US-0131449.	
PR	30-APR-1999; 99US-0132048.	
PR	30-APR-1999; 99US-0132407.	
PR	04-MAY-1999; 99US-0132484.	
PR	05-MAY-1999; 99US-0132485.	
PR	06-MAY-1999; 99US-0132486.	
PR	06-MAY-1999; 99US-0132487.	
PR	07-MAY-1999; 99US-0132863.	
PR	11-MAY-1999; 99US-0134256.	
PR	14-MAY-1999; 99US-0134218.	
PR	14-MAY-1999; 99US-0134219.	
PR	14-MAY-1999; 99US-0134221.	
PR	14-MAY-1999; 99US-0134370.	
PR	18-MAY-1999; 99US-0134768.	
PR	19-MAY-1999; 99US-0134941.	
PR	20-MAY-1999; 99US-0135124.	
PR	21-MAY-1999; 99US-0135353.	
PR	24-MAY-1999; 99US-0135629.	
PR	25-MAY-1999; 99US-0136021.	
PR	27-MAY-1999; 99US-0136392.	
PR	28-MAY-1999; 99US-0136782.	
PR	01-JUN-1999; 99US-0137222.	
PR	03-JUN-1999; 99US-0137528.	
PR	04-JUN-1999; 99US-0137502.	

PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144085.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 06-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.

PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159299.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 28-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 9.6%; Score 296.5; DB 21; Length 532;
Best Local Similarity 22.7%; Pred. No. 1.2e-17;
Matches 128; Conservative 96; Mismatches 235; Indels 105; Gaps 19;

RESULT 9

ID AAG11240 standard; Protein; 505 AA.
XX
AC AAG11240;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 9883.
XX
XX Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
XX 06-SEP-2000.
PD
XX 25-FEB-2000; 2000EP-0301439.
FF
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0128845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137724.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149923.
PR 23-AUG-1999; 99US-0149902.

PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 14-JUN-1999; 99US-0138847.
PR 10-AUG-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 08-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149388.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150586.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 23-OCT-1999; 99US-0161408.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.

PR	23-MAR-1999	99US-0125789
PR	24-MAR-1999	99US-0125790
PR	25-MAR-1999	99US-0126264
PR	26-MAR-1999	99US-0126785
PR	27-MAR-1999	99US-0127462
PR	01-APR-1999	99US-0128234
PR	06-APR-1999	99US-0128714
PR	08-APR-1999	99US-0129845
PR	16-APR-1999	99US-0130077
PR	19-APR-1999	99US-0130049
PR	21-APR-1999	99US-0130510
PR	23-APR-1999	99US-0130691
PR	23-APR-1999	99US-0131449
PR	28-APR-1999	99US-0132048
PR	30-APR-1999	99US-0132307
PR	30-APR-1999	99US-0133407
PR	04-MAY-1999	99US-0133484
PR	05-MAY-1999	99US-0133485
PR	06-MAY-1999	99US-0132486
PR	06-MAY-1999	99US-0132487
PR	07-MAY-1999	99US-0133863
PR	11-MAY-1999	99US-0134256
PR	14-MAY-1999	99US-0134218
PR	14-MAY-1999	99US-0134219
PR	14-MAY-1999	99US-0134221
PR	14-MAY-1999	99US-0134370
PR	18-MAY-1999	99US-0134768
PR	19-MAY-1999	99US-0134941
PR	20-MAY-1999	99US-0135124
PR	21-MAY-1999	99US-0135353
PR	24-MAY-1999	99US-0135629
PR	25-MAY-1999	99US-0136021
PR	27-MAY-1999	99US-0136392
PR	28-MAY-1999	99US-0136782
PR	01-JUN-1999	99US-0137222
PR	03-JUN-1999	99US-0137528
PR	04-JUN-1999	99US-0137502
PR	07-JUN-1999	99US-0137724
PR	08-JUN-1999	99US-0138094
PR	10-JUN-1999	99US-0138540
PR	10-JUN-1999	99US-0138847
PR	14-JUN-1999	99US-0139119
PR	16-JUN-1999	99US-0139452
PR	16-JUN-1999	99US-0139453
PR	17-JUN-1999	99US-0139492
PR	18-JUN-1999	99US-0139454
PR	18-JUN-1999	99US-0139455
PR	18-JUN-1999	99US-0139456
PR	18-JUN-1999	99US-0139750
PR	18-JUN-1999	99US-0139763
PR	21-JUN-1999	99US-0139817
PR	22-JUN-1999	99US-0139899
PR	23-JUN-1999	99US-0140353
PR	23-JUN-1999	99US-0140354
PR	24-JUN-1999	99US-0139461
PR	18-JUN-1999	99US-0139462
PR	18-JUN-1999	99US-0139463
PR	18-JUN-1999	99US-0139750
PR	18-JUN-1999	99US-0139763
PR	21-JUN-1999	99US-0139817
PR	22-JUN-1999	99US-0139899
PR	23-JUN-1999	99US-0140353
PR	23-JUN-1999	99US-0140354
PR	24-JUN-1999	99US-0139461
PR	24-JUN-1999	99US-0139462
PR	24-JUN-1999	99US-0139463
PR	28-JUN-1999	99US-0140823
PR	29-JUN-1999	99US-0140991
PR	30-JUN-1999	99US-0141287
PR	01-JUL-1999	99US-0141842
PR	01-JUL-1999	99US-0142154
PR	02-JUL-1999	99US-0142055
PR	06-JUL-1999	99US-0142390
PR	08-JUL-1999	99US-0142803
PR	09-JUL-1999	99US-0142920
PR	12-JUL-1999	99US-0142977
PR	13-JUL-1999	99US-0143542
PR	14-JUL-1999	99US-0143624
PR	15-JUL-1999	99US-0144005

PR	16-JUL-1999;	99US-0144085.	PR	07-OCT-1999;	99US-0158029.
PR	16-JUL-1999;	99US-0144086.	PR	08-OCT-1999;	99US-0158212.
PR	19-JUL-1999;	99US-0144325.	PR	12-OCT-1999;	99US-0158369.
PR	19-JUL-1999;	99US-0144331.	PR	13-OCT-1999;	99US-0159293.
PR	19-JUL-1999;	99US-0144332.	PR	13-OCT-1999;	99US-0159294.
PR	19-JUL-1999;	99US-0144333.	PR	13-OCT-1999;	99US-0159295.
PR	19-JUL-1999;	99US-0144334.	PR	14-OCT-1999;	99US-0159329.
PR	19-JUL-1999;	99US-0144335.	PR	14-OCT-1999;	99US-0159330.
PR	20-JUL-1999;	99US-0144352.	PR	14-OCT-1999;	99US-0159331.
PR	20-JUL-1999;	99US-0144632.	PR	14-OCT-1999;	99US-0159637.
PR	20-JUL-1999;	99US-0144884.	PR	14-OCT-1999;	99US-0159638.
PR	21-JUL-1999;	99US-0144814.	PR	18-OCT-1999;	99US-0159584.
PR	21-JUL-1999;	99US-0145086.	PR	21-OCT-1999;	99US-0160741.
PR	21-JUL-1999;	99US-0145088.	PR	21-OCT-1999;	99US-0160767.
PR	22-JUL-1999;	99US-0145085.	PR	21-OCT-1999;	99US-0160768.
PR	22-JUL-1999;	99US-0145087.	PR	21-OCT-1999;	99US-0160770.
PR	22-JUL-1999;	99US-0145089.	PR	21-OCT-1999;	99US-0160814.
PR	22-JUL-1999;	99US-0145192.	PR	21-OCT-1999;	99US-0160815.
PR	23-JUL-1999;	99US-0145145.	PR	22-OCT-1999;	99US-0160980.
PR	23-JUL-1999;	99US-0145218.	PR	22-OCT-1999;	99US-0160981.
PR	23-JUL-1999;	99US-0145224.	PR	22-OCT-1999;	99US-0160989.
PR	26-JUL-1999;	99US-0145276.	PR	25-OCT-1999;	99US-0161404.
PR	27-JUL-1999;	99US-0145913.	PR	25-OCT-1999;	99US-0161405.
PR	27-JUL-1999;	99US-0145918.	PR	25-OCT-1999;	99US-0161406.
PR	27-JUL-1999;	99US-0145919.	PR	26-OCT-1999;	99US-0161359.
PR	28-JUL-1999;	99US-0145951.	PR	26-OCT-1999;	99US-0161360.
PR	02-AUG-1999;	99US-0146386.	PR	26-OCT-1999;	99US-0161361.
PR	02-AUG-1999;	99US-0146388.	PR	28-OCT-1999;	99US-0161920.
PR	02-AUG-1999;	99US-0146389.	PR	28-OCT-1999;	99US-0161922.
PR	03-AUG-1999;	99US-0147038.	PR	28-OCT-1999;	99US-0161993.
PR	04-AUG-1999;	99US-0147204.	PR	29-OCT-1999;	99US-0162142.
PR	04-AUG-1999;	99US-0147302.			
PR	05-AUG-1999;	99US-0147192.			
PR	05-AUG-1999;	99US-0147260.			
PR	06-AUG-1999;	99US-0147303.			
PR	06-AUG-1999;	99US-0147416.			
PR	09-AUG-1999;	99US-0147493.			
PR	09-AUG-1999;	99US-0147935.			
PR	10-AUG-1999;	99US-0148171.			
PR	11-AUG-1999;	99US-0148319.			
PR	12-AUG-1999;	99US-0148341.			
PR	13-AUG-1999;	99US-0148565.			
PR	13-AUG-1999;	99US-0148684.			
PR	16-AUG-1999;	99US-0149368.			
PR	17-AUG-1999;	99US-0149175.			
PR	18-AUG-1999;	99US-0149426.			
PR	20-AUG-1999;	99US-0149722.			
PR	20-AUG-1999;	99US-0149723.			
PR	20-AUG-1999;	99US-0149929.			
PR	23-AUG-1999;	99US-0149902.			
PR	23-AUG-1999;	99US-0149930.			
PR	25-AUG-1999;	99US-0150566.			
PR	26-AUG-1999;	99US-0150884.			
PR	27-AUG-1999;	99US-0151065.			
PR	27-AUG-1999;	99US-0151066.			
PR	27-AUG-1999;	99US-0151080.			
PR	30-AUG-1999;	99US-0151303.			
PR	31-AUG-1999;	99US-0151438.			
PR	01-SEP-1999;	99US-0151930.			
PR	07-SEP-1999;	99US-0152363.			
PR	10-SEP-1999;	99US-0153070.			
PR	13-SEP-1999;	99US-0153758.			
PR	15-SEP-1999;	99US-0154018.			
PR	16-SEP-1999;	99US-0154039.			
PR	20-SEP-1999;	99US-0154779.			
PR	22-SEP-1999;	99US-0155139.			
PR	23-SEP-1999;	99US-0155486.			
PR	24-SEP-1999;	99US-0155659.			
PR	28-SEP-1999;	99US-0156458.			
PR	29-SEP-1999;	99US-0156596.			
PR	04-OCT-1999;	99US-0157117.			
PR	05-OCT-1999;	99US-0157753.			
PR	06-OCT-1999;	99US-0157865.			
Query Match 9.1%; Score 282; DB 21; Length 532;			Best Local Similarity 20.8%; Pred. No. 2.5e-16;		
Matches 126; Conservative 98; Mismatches 224; Indels 158; Gaps 21;					
QY	49	OSKOPQFLKSIDELAAFSVAVETPKRO-----FDDLOKHIESIENAIKSLKESNG 98	QY	49	OSKOPQFLKSIDELAAFSVAVETPKRO-----FDDLOKHIESIENAIKSLKESNG 98
DB	12	KSSQSPFPE-----FOKQASLMTSCNLLWKELSEHFTSMQNLMMKKSEALR 57	DB	12	KSSQSPFPE-----FOKQASLMTSCNLLWKELSEHFTSMQNLMMKKSEALR 57
QY	99	VLAARNNNHQPMLSPRRNNVSVEITTVVS-----OPQOEIVPETSNDP 143	QY	99	VLAARNNNHQPMLSPRRNNVSVEITTVVS-----OPQOEIVPETSNDP 143
DB	58	QMIEITDNTQTSIELLKREVTIDHSVEIAGKVEERVAALDSLEKARDGDEDTGEV 117	DB	58	QMIEITDNTQTSIELLKREVTIDHSVEIAGKVEERVAALDSLEKARDGDEDTGEV 117
QY	144	EGG-----RMCELMCSKGLRKYIYANISDQAKLMEIPPSALKLAKPAKVLDCIGK 195	QY	144	EGG-----RMCELMCSKGLRKYIYANISDQAKLMEIPPSALKLAKPAKVLDCIGK 195
DB	118	DDGGLLSALKSLCKMDARGFWGFIARKKELENRSQIPVALVDCVPPKLVLEAVE 177	DB	118	DDGGLLSALKSLCKMDARGFWGFIARKKELENRSQIPVALVDCVPPKLVLEAVE 177
QY	196	FYLOGRRAFTKESPMSSARQVSLILLESF---LMPDRGKGKVKIESWIKDEBETAANAV 252	QY	196	FYLOGRRAFTKESPMSSARQVSLILLESF---LMPDRGKGKVKIESWIKDEBETAANAV 252
DB	178	VFPVDKRG-GGEKVSNDFGWACVVILESLIPVMVDPVMGKSRLLVTPSVKERAKETAETW 236	DB	178	VFPVDKRG-GGEKVSNDFGWACVVILESLIPVMVDPVMGKSRLLVTPSVKERAKETAETW 236
QY	253	RKRLMTEGLAAAEKMDARGILLLVACFGVPSNFRSTDLDLIRMSGNEIAGALKRSQF 312	QY	253	RKRLMTEGLAAAEKMDARGILLLVACFGVPSNFRSTDLDLIRMSGNEIAGALKRSQF 312
DB	237	KASLEERGIEENVKTPDVHTFLOHLVTFGIVKK-----DDJALYR-----KLIVGSAWRKOM 288	DB	237	KASLEERGIEENVKTPDVHTFLOHLVTFGIVKK-----DDJALYR-----KLIVGSAWRKOM 288
QY	313	-----LVPWVSIGIVESITKGMHIEALEMVYTFGEMDKFSAALVLTSLFKMSK----- 360	QY	313	-----LVPWVSIGIVESITKGMHIEALEMVYTFGEMDKFSAALVLTSLFKMSK----- 360
DB	289	PKLAVSVGLGQMPDMEIELIIRGQQLDAVHFTFVGLVHLFPPVPLLKAYLRDAKKATA 348	DB	289	PKLAVSVGLGQMPDMEIELIIRGQQLDAVHFTFVGLVHLFPPVPLLKAYLRDAKKATA 348
QY	361	---ESFERAKKAQSPKAPKEATKQLAVLSVYQCMETHKLDPAKELPGWQIKQIVSL 417	QY	361	---ESFERAKKAQSPKAPKEATKQLAVLSVYQCMETHKLDPAKELPGWQIKQIVSL 417
DB	349	LITDDSNNSGRSAHL-----VARKEQSALRAVLKICEEYKLE--EEFPPENLKGLDQL 400	DB	349	LITDDSNNSGRSAHL-----VARKEQSALRAVLKICEEYKLE--EEFPPENLKGLDQL 400
QY	418	EKDTQLDLDKEMEERKARSLSLMEBAALAKRMVNOQIKRPLRS-PMEMPPVTSSSYSDIYD 476	QY	418	EKDTQLDLDKEMEERKARSLSLMEBAALAKRMVNOQIKRPLRS-PMEMPPVTSSSYSDIYD 476
DB	401	EKTTEKEKPAVIPAN-----KXTRASYSYSGMPMPAKAGRIT----- 436	DB	401	EKTTEKEKPAVIPAN-----KXTRASYSYSGMPMPAKAGRIT----- 436
QY	477	RSFPQRDDDDQDEISALVSYLGPSTSFPHRSRRSPBYMVPLPHGGLGRSVAYEHLAPN 536	QY	477	RSFPQRDDDDQDEISALVSYLGPSTSFPHRSRRSPBYMVPLPHGGLGRSVAYEHLAPN 536
DB	437	-----NAYVSSFPFPPPTFIRSQSHSPQYGVF-----AYTTSPT 471	DB	437	-----NAYVSSFPFPPPTFIRSQSHSPQYGVF-----AYTTSPT 471
QY	537	SY---SPGHGHLRHQVSPSLVHG--QRHPLQY-----SPEI-HGQQOLPY 576	QY	537	SY---SPGHGHLRHQVSPSLVHG--QRHPLQY-----SPEI-HGQQOLPY 576

```
Db      | 472 IYSNRSPY-----QYSPEAVHGSYQTSVSYTAYGTCVSPAAPPVYHPHPHHH 525
Qy      | 577 GIORVY 582
Db      | 526 HIQHAY 531
RESULT 13
AAG11241
ID AAG11241 standard; Protein; 483 AA.
AC AAG11241;
XX
XX 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 9884.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
FN EPI033405-A2.
PD 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127482.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134219.
PR 18-MAY-1999; 99US-0134370.
PR 19-MAY-1999; 99US-0134768.
PR 21-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142377.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
```

PR 13-AUG-1999; 99US-0148341.
 PR 13-AUG-1999; 99US-0148565.
 PR 13-AUG-1999; 99US-0148684.
 PR 16-AUG-1999; 99US-0149368.
 PR 17-AUG-1999; 99US-0149175.
 PR 18-AUG-1999; 99US-0149426.
 PR 20-AUG-1999; 99US-0149722.
 PR 20-AUG-1999; 99US-0149723.
 PR 20-AUG-1999; 99US-0149929.
 PR 23-AUG-1999; 99US-0149902.
 PR 23-AUG-1999; 99US-0149930.
 PR 25-AUG-1999; 99US-0150566.
 PR 26-AUG-1999; 99US-0150884.
 PR 27-AUG-1999; 99US-0151065.
 PR 27-AUG-1999; 99US-0151066.
 PR 27-AUG-1999; 99US-0151080.
 PR 30-AUG-1999; 99US-0151303.
 PR 31-AUG-1999; 99US-0151438.
 PR 01-SEP-1999; 99US-0151930.
 PR 07-SEP-1999; 99US-0152363.
 PR 10-SEP-1999; 99US-0153070.
 PR 13-SEP-1999; 99US-0153758.
 PR 15-SEP-1999; 99US-0154018.
 PR 16-SEP-1999; 99US-0154039.
 PR 20-SEP-1999; 99US-0154779.
 PR 22-SEP-1999; 99US-0155139.
 PR 23-SEP-1999; 99US-0155486.
 PR 24-SEP-1999; 99US-0155659.
 PR 28-SEP-1999; 99US-0156458.
 PR 29-SEP-1999; 99US-0156596.
 PR 04-OCT-1999; 99US-0157117.
 PR 05-OCT-1999; 99US-0157753.
 PR 06-OCT-1999; 99US-0157865.
 PR 07-OCT-1999; 99US-0158029.
 PR 08-OCT-1999; 99US-0158232.
 PR 12-OCT-1999; 99US-0158369.
 PR 13-OCT-1999; 99US-0159293.
 PR 13-OCT-1999; 99US-0159294.
 PR 13-OCT-1999; 99US-0159295.
 PR 14-OCT-1999; 99US-0159329.
 PR 14-OCT-1999; 99US-0159330.
 PR 14-OCT-1999; 99US-0159331.
 PR 14-OCT-1999; 99US-0159637.
 PR 14-OCT-1999; 99US-0159638.
 PR 18-OCT-1999; 99US-0159584.
 PR 21-OCT-1999; 99US-0160741.
 PR 21-OCT-1999; 99US-0160767.
 PR 21-OCT-1999; 99US-0160768.
 PR 21-OCT-1999; 99US-0160770.
 PR 21-OCT-1999; 99US-0160814.
 PR 21-OCT-1999; 99US-0160815.
 PR 22-OCT-1999; 99US-0160980.
 PR 22-OCT-1999; 99US-0160981.
 PR 22-OCT-1999; 99US-0160989.
 PR 25-OCT-1999; 99US-0161404.
 PR 25-OCT-1999; 99US-0161405.
 PR 25-OCT-1999; 99US-0161406.
 PR 25-OCT-1999; 99US-0161359.
 PR 26-OCT-1999; 99US-0161360.
 PR 26-OCT-1999; 99US-0161361.
 PR 28-OCT-1999; 99US-0161920.
 PR 28-OCT-1999; 99US-0161992.
 PR 28-OCT-1999; 99US-0161993.
 PR 29-OCT-1999; 99US-0162142.

Query Match 9.1%; Score 281.5; DB 21; Length 483;
 Best Local Similarity 22.4%; Pred. No. 2.4e-16;
 Matches 121; Conservative 92; Mismatches 215; Indels 111; Gaps 18;
 74 RQFDDLQKIESNAIDSKLESNGVLAARNNFHQPMLSPRRNVSVETTVV-----128
 2 KKSEALKQMIETLDNQTQTSLE-----SLKREVTIDHSVIVAGVKV 43

QY 129 -----SQPSQEIIVPETSINKPEGG-----RMCELMCSKGLRXYIYANIS 166
 DB 44 GERAPAALESLEKARDGCGSNDSDGVDDEGLSALKSLCLKMDARGFWMFVTKRK 103
 QY 167 DOAKLMEIIPSAKLAKPAKFLVDCIGKFYLOGRRAFTKESPMSSARQVSLILLESFLL 226
 DB 104 ELENLSKIPALVDCVDPAMLVLEALISEVFPVDTGRDKVSYNDYGWACVVILSLTPVIV 163
 QY 227 MPDRGKGVKIESWIKOEAETAANVWRKRLMTBEGGLAAAEKMDARGLLLLVACFGVPSNF 286
 DB 164 DEVIGKSLLVTPSVKKEKAEIAETWKKSLEERGRIENVKTPDVHTFLOHLVTFGI---V 220
 QY 287 RSTDLDLIRMSGNEIAGALKRSOP-----LVPWVGIVESSIKRGMHIEALEMVY 338
 DB 221 KSEDALYRKL-----VVGSAWRKQMPKLAVSVGLGDQMPDMIEELISRCQQLDAVHTY 275
 QY 339 TFGMEDKFSALVLTSLFKMSKESFER-AKRKAQSPFLAFKEATKQJLAVLSSVMQCMETH 397
 DB 276 EVGLVDKPPVPLKAVLRDAKSAASIMEDSSNTGRTATHLVARKEQSALKAVLKCIIEY 335
 QY 398 KLDPAKELPGWQIKQIVSLEKDTLQLDKEMEERKARSLSLMEBAALAKRMVNOQIKRPRL 457
 DB 336 KLE--EEFPPENLKRRLDQLEKTKTKRKAAPVAN-----KRTA 374
 QY 458 S---PMEMPVTSSSYPIYDRSPFQRRDDQDEISALVSSYLGPTSTPFHRSRSPY 514
 DB 375 SYNGP--MPPAKAGRITNAVYS-SFTIRSPSHSPQVAPAPVSPPTTV--YSNRSPY 429
 QY 515 -----WVPLPHGGLGRSVYAYEHL-----APNSYSPGCHGLHRLHQVSPSLVHGQRH 560
 DB 430 PYSPEIIFSGYQSGPIGYPAVNGYCVGFPAPAPVFPVYHP--HHQF--HHQF 481

RESULT 14
 AAG49609
 ID AAG49609 standard; Protein; 483 AA.
 XX AAG49609;
 AC AAG49609;
 XX
 DT 18-OCT-2000 (first entry)
 XX
 DE Arabidopsis thaliana protein fragment SEQ ID NO: 62775.
 XX
 KW Protein identification; signal transduction pathway; metabolic pathway;
 KW hybridisation assay; genetic mapping; gene expression control; promoter;
 KW termination sequence.
 XX
 OS Arabidopsis thaliana.
 XX
 PN EP1033405-A2.
 XX
 PD 06-SEP-2000.
 XX
 PF 25-FEB-2000; 2000EP-0301439.
 XX
 PR 25-FEB-1999; 99US-0121825.
 PR 05-MAR-1999; 99US-0123180.
 PR 09-MAR-1999; 99US-0123548.
 PR 23-MAR-1999; 99US-0125788.
 PR 25-MAR-1999; 99US-0126264.
 PR 29-MAR-1999; 99US-0126785.
 PR 01-APR-1999; 99US-0127462.
 PR 06-APR-1999; 99US-0128234.
 PR 08-APR-1999; 99US-0128714.
 PR 16-APR-1999; 99US-0128845.
 PR 19-APR-1999; 99US-0130077.
 PR 21-APR-1999; 99US-0130449.
 PR 23-APR-1999; 99US-0130510.
 PR 23-APR-1999; 99US-0130891.
 PR 28-APR-1999; 99US-0131449.
 PR 30-APR-1999; 99US-0132048.
 PR 30-APR-1999; 99US-0132407.
 PR 04-MAY-1999; 99US-0132484.

PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160880.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 26-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.
PR 29-OCT-1999; 99US-0162142.
Query Match 9.1%; Score 281.5; DB 21; Length 483;
Best Local Similarity 22.4%; Pred. No. 2.4e-16;
Matches 121; Conservative 92; Mismatches 215; Indels 111; Gaps 18;
QY 74 ROPDDLOKHIESIENAIKSLKNSGVLAARNNNPHQMLSPRRNNVSVETTVV----- 128
Db 2 KXSEALKQMIETLDNQTSLE-----SLKREVTIDHSVEIVAGKV 43
QY 129 -----SPSQEIVPETNKPEGG-----RMCELMCKSLKRYIVANIS 166
Db 44 GERARALESLEKARDGCGDGNDSGDVDEEGLLSALKSLCKMDARGFWNFVARKK 103
QY 167 DQAKLMEETPSALKLAKEPAKFLVLCIGKFLVQGRRAFTKESPMSSARQVSLILLESFLL 226
Db 104 ELENLSRKIPALVDCVDFAMLVLEAISEVFPVDRGKVDNDYGCWACVILESLTPVIV 163
QY 227 MDRGKGKVKIESWKDEAETAAVWRKMLMTEGGLAAEKWDARGILLLVACFGVSNF 286
Db 164 DPVIGKRLVTPSVKBEKAKETAETWKKSLERGIENVKTPDVHTFLOHLVTFGI---V 220
QY 287 RSTDLLDLRMSSNEIAGALKRSQF-----LVPWVSGIVRESSIKRGMHIEALEMVI 338
Db 221 KSEDALYKRL-----VUGSANRQKMPKLAVSGLDQMPDMIEELISRGQOLDVAHFY 275
QY 339 TFGMEDKPSAALVLTSLFKWSEFER-AKRAKAQSLAFKAAATKQLAVLSSVNCMETH 397
Db 276 EVGLVDKFPFVPLLKAYLRDAKSAASIMEDSSNTGRATHLVARKEQALKAVLKCEEY 335
QY 398 KLPAPAKELQWQKEOIVLSLEKDTLQDKEMEKEKARSLSLMEERAAKRMVNOQIKPRL 457
Db 336 KLE--EEFPENLKRLDQLEKTEKRPAAVPA-----KXTRA 374
QY 458 S---PMEMPVTSSSYPIYRDRSPSQRDDQDEISALVSYLGPSTSPFHRGRRSPY 514
Db 375 SYNGP--MPPAKAGRITNAYVS--SPFFIRSPSHSPQYASPAAYSPPTTV--YSNRSPPY 429
QY 515 -----WVPLPHGGLGRSVAYEHL-----APNSYSPGHGHLHRCYPSLVHGQPH 560
Db 430 PYSPEIIPSGYQSPGYPAYNGYCNVPAPAPVYHP--HHQ-----HHQOH 481
RESULT 15
AAG19754
ID AAG19754 standard; Protein; 505 AA.
XX
AC AAG19754;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 21674.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.

XX EP1033405-A2.
XX 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127482.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.

PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149923.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 13-SEP-1999; 99US-0153070.
PR 15-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.

PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161320.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 8.9%; Score 277; DB 21; Length 505;
Best Local Similarity 21.6%; Pred. No. 6.5e-16;
Matches 122; Conservative 94; Mismatches 226; Indels 124; Gaps 20;

QY 65 FSVAVETFKQDDLOKHIESIENAIIDSKLENGVYLAARNNF-HQWMLSPRRNVSV 123
DB 15 FTSMEQNLKKSEALRQMIETLNDQSSIE---LLKHREVTIDHSVEIAEGKVEERV 70
QY 124 TTVTVSQSQEIVPETSINKPEGG-----RMCELMCSKGLRKYYIYANISDQAKLMEEI 175
DB 71 AALDSLEKARDCGDEDTGEVDDGDGLLSALKSLCLKMDARGFWGVIAKKLENLRSQI 130
QY 176 PSALKLAKEPAKFVLDICIGFYLGRRFTKESPMSSARQVSLILIESF---LLMPDRGK 232
DB 131 PVALVDCVDPKLVLEAVSEVPVDKRG-GGEKVSNDFCWACVWILESLIPVWDPVWGK 189
QY 233 GKVKIESWIKDEAETAFAVAKRELMTTEGLAAEKDARGILLLVACFCVPSNFRSTOLL 292
DB 190 SRLLVTPSVKEKAKETATWKASLEERGGENVKTDPVHTFLOHLVTFGVKK---DDL 245
QY 293 DLIRMSGSEIAGALKRSQF-----LVPVMSGIVESIKRGMHIEALEMVTYFGMED 344
DB 246 ALYR----KLAVGSANRQMPKLVASVGLGDQDQPMIEELIIRGQQLDAVHTFEVGLVH 301
QY 345 KFSAAVLVTSFLKMSK-----ESFERAKRKAQSPKAFKEAATQQLAVLSSVMQCMETH 397
DB 302 LFPEVPLLKAYLRDAKKATALITDDSNNSGRSAHL-----VARKEQSALRAVLKICIEY 355
QY 398 KLDPAKELPGWQIKEOIVSLEKDTLOLDKEMEKEKARSLSIMEEALAKRMVNOQIKRPL 457
DB 356 KLE--EEFPENLKXLDQLEKTKRKRPVIPAN-----KRTPA 394

QY	458	S-PMENPPVTSSSYSPFIYDRSPSQRDDQDEISALVSSYLGPSTSPPHRSRSPYMV	516
Db	395	SYSGPMPPAXAGRIT-----NAYVSSFPPTFIRSQSHSPQYGV	435
QY	517	PLPHGGLGRSVYAEHLAPNSY---SPGHGRLHRQYSPSLVHG--ORHPLQY-----	564
Db	436	P-----AYTSPPTIYSNRSPY-----QYSPEAVHGSYQTSFVSYETAYGTY	478
QY	565	-----SPPI-HGQQQLPYGIQRYV	582
Db	479	CSFVAAPPPVYHPHHHHHICHAY	504

Search completed: August 14, 2003, 10:37:33
Job time : 88 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 14, 2003, 10:36:04 ; Search time 29 seconds
(without alignments)
888,528 Million cell updates/sec

Title: US-09-890-475-1

Perfect score: 3104

Sequence: 1 MSNYPTVAQAQPTTTANPL.....RYLGLNQRSPNSLDPK 609

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: /cgn2_6/prodata/1/iaa/5A_COMB.pep.*
- 2: /cgn2_6/prodata/1/iaa/5B_COMB.pep.*
- 3: /cgn2_6/prodata/1/iaa/5A_COMB.pep.*
- 4: /cgn2_6/prodata/1/iaa/5B_COMB.pep.*
- 5: /cgn2_6/prodata/1/iaa/5A_COMB.pep.*
- 6: /cgn2_6/prodata/1/iaa/5B_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match %	Score	Length	ID	Description
1	143.5	4.6	1201	3	US-09-098-901-2 Sequence 2, Appli
2	137.5	4.4	2074	4	US-09-491-356C-9 Sequence 9, Appli
3	123.5	4.0	2023	4	US-09-491-356C-8 Sequence 8, Appli
4	122.5	3.9	2285	4	US-09-308-375-2 Sequence 2, Appli
5	120.5	3.9	1001	3	US-09-060-410-2 Sequence 2, Appli
6	120.5	3.9	1001	4	US-09-723-458-2 Sequence 2, Appli
7	119.5	3.8	1388	2	US-08-685-576-1 Sequence 1, Appli
8	117.5	3.8	828	1	US-08-261-304-2 Sequence 2, Appli
9	113.5	3.7	816	2	US-08-533-306A-6 Sequence 6, Appli
10	113.5	3.7	816	2	US-08-742-922A-6 Sequence 6, Appli
11	113.5	3.7	885	2	US-08-533-306A-4 Sequence 4, Appli
12	113.5	3.7	885	2	US-08-742-923A-4 Sequence 4, Appli
13	112	3.6	3248	1	US-08-353-700-1 Sequence 1, Appli
14	112	3.6	3248	5	PCT-US95-16216-1 Sequence 1, Appli
15	110	3.5	2482	1	US-08-328-254-6 Sequence 6, Appli
16	109	3.5	578	3	US-09-066-046-6 Sequence 6, Appli
17	109	3.5	578	3	US-08-375-762-50 Sequence 50, Appli
18	109	3.5	578	3	US-09-295-028-50 Sequence 50, Appli
19	109	3.5	578	4	US-09-106-582-50 Sequence 36, Appli
20	109	3.5	1174	2	US-08-446-145-36 Sequence 36, Appli
21	108.5	3.5	803	4	US-09-252-991A-28407 Sequence 28407, A
22	108.5	3.5	1093	5	PCT-US93-03077-1 Sequence 1, Appli
23	107.5	3.5	736	4	US-09-252-991A-21498 Sequence 21498, A
24	106	3.4	1618	1	US-07-853-913-4 Sequence 4, Appli
25	106	3.4	2293	3	US-09-368-590-2 Sequence 2, Appli
26	105	3.4	935	4	US-09-914-259-25 Sequence 25, Appli
27	104.5	3.4	784	4	US-09-740-235-1 Sequence 1, Appli

28	104.5	3.4	1075	2	US-08-993-228-19 Sequence 19, Appli
29	104.5	3.4	1196	1	US-08-144-121-4 Sequence 4, Appli
30	104.5	3.4	1196	2	US-08-735-893-4 Sequence 4, Appli
31	104.5	3.4	1786	3	US-08-973-462-8 Sequence 8, Appli
32	104	3.4	1886	4	US-08-938-105-3 Sequence 3, Appli
33	104	3.4	1939	4	US-09-310-187A-1 Sequence 1, Appli
34	103.5	3.3	749	1	US-08-369-796-8 Sequence 8, Appli
35	103.5	3.3	749	2	US-08-852-091-8 Sequence 8, Appli
36	103.5	3.3	749	3	US-08-820-754-8 Sequence 8, Appli
37	103.5	3.3	749	3	US-08-956-652-8 Sequence 8, Appli
38	103.5	3.3	749	3	US-08-956-869-8 Sequence 8, Appli
39	103.5	3.3	749	3	US-09-012-710-7 Sequence 7, Appli
40	103.5	3.3	749	3	US-08-948-547-8 Sequence 8, Appli
41	103.5	3.3	749	4	US-09-556-273-7 Sequence 7, Appli
42	103.5	3.3	749	4	US-08-956-653A-8 Sequence 8, Appli
43	103.5	3.3	749	5	PCT-US95-17025-8 Sequence 8, Appli
44	102.5	3.3	737	1	US-08-188-582-16 Sequence 16, Appli
45	102.5	3.3	737	1	US-08-646-715-16 Sequence 16, Appli

ALIGNMENTS

RESULT 1

US-09-098-901-2
; Sequence 2, Application US/09098901B
; Patent No. 6218144
; GENERAL INFORMATION:
; APPLICANT: Scott, Matthew
; APPLICANT: Sisson, John C.
; TITLE OF INVENTION: Costal2 Genes and their Uses
; FILE REFERENCE: SUN-65P
; CURRENT APPLICATION NUMBER: US/09/098,901B
; CURRENT FILING DATE: 1998-06-17
; EARLIER APPLICATION NUMBER: 60/051,347
; EARLIER FILING DATE: 1997-06-30
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1201
; TYPE: PRT
; ORGANISM: D. Melanogaster
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(1201)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-09-098-901-2

Query Match 4.6%; Score 143.5; DB 3; Length 1201;
Best Local Similarity 19.9%; Pred. No. 0.00011;
Matches 133; Conservative 100; Mismatches 232; Indels 203; Gaps 30;

Qy	18	PLIQRHQSQRRELPKIVETESTMDITIGSKQOPFLKSIDELAAPSVAVETFKQFD	77
Db	498	FILEABEPEDVNSEANSESPNSDNDTNDNESHHPDLDKIESL-----MEBFROKTD	551
Qy	78	D-----LOKHESIEINAIDSKLES-----NG-----VVLAARNNNFHOMLSPRRNN	119
Db	552	ALLLEKHAELVSKHPKAVMQSDRIEIAQPEENGDDKVKVIGSRRSV-OPGASLSTAE	610
Qy	120	VSVETTVTVQSQSIIVPET-----SNKPEGRCMELMCS-----KGLRKYIYA	163
Db	611	LAMLNKVASQCPFPIDPESVVDPLESSGEGIRQAALAAAAATAPIBQLQKLRKLV-A	669
Qy	164	NISDOAKLMEEIPSAKLA-----KEPAKFLVDC-----IGK	195
Db	670	EIEGKORQLREIEEIQVKNIIELVKNSDTRSHAKORPHKRAKLEACDCKAKQLGK	729
Qy	196	FYIQR-----RAFTKESPM-----SSARQVSL-----ILSFILMPDRG	231
Db	730	ALVQGRGQSEIERWTITIGHLERRLEDLSMKHIAGESGQVKVKKLQOSVGSERKQADLQ	789
Qy	232	KGKVKIESIKDEAETAAVAKRLMTGGLAAAE---KMDARGILLVAVCFGVPSNFRS	288

```
Db 790 K-KRKECKLRQMEAEVLKRESRETKELVKAGQGPQOQRLKAVQA-----RI 840
QY 289 TDLILLRMSGN-RIAGALKRSQPL---VPMVSGIVESSIKRGMH-----330
Db 841 THLNHILREKSNLXQPGPQOQETLRHEIRNLRTGRLDLLKERCHELDKRLKRDKVLTQK 900
QY 331 -----IEALEMVTYTFGMEKDFSAALVLTSLFKMSKESPERAKRKAQSPFLAFKEA 379
Db 901 BERRKLECEAEIDAAEIFNEMITCHRSIDTSD-RIQREKGEQMLMARLNLSTEEM 959
QY 380 ATKQLAVLSSVMQC-----METHKLDPAKELPGWQIKQEIIVSL-----DKTLLQ 424
Db 960 RTLLYKYFTKVDLRDSRKLEQLVQLERERDAWEMKERVLSNAVQAQALEGERNVALL 1019
QY 425 DKEMBEKARSLSM-----EEAALAKRMYNQOIKRPRLSPMEMPPVTSSS---YSPITYRD 476
Db 1020 ORCHEMK---LTLMLRHAETESSASSYGERALAP---ACVAPVQASDDFYDHFYKG 1073
QY 477 RSFPQORDDQDEISALVSSYLGPSSTSPHRSRSPSPVMDPLPHGGL-----GR 525
Db 1074 GGNPSK-----ALT-----KAPK---PMPTGSALDKYKQKORSGR 1106
QY 526 SVYAYEHL 533
Db 1107 NIFAKFHV 1114

RESULT 2
US-09-491-356C-9
; Sequence 9, Application US/09491356C
; Patent No. 6566061
; GENERAL INFORMATION:
; APPLICANT: Philibert, Robert A.
; APPLICANT: Ginns, Edward I.
; TITLE OF INVENTION: IDENTIFICATION OF POLYMORPHISMS IN THE PCTG4 REGION OF XQ13
; FILE REFERENCE: 9465.6US11
; CURRENT FILING DATE: 2000-01-26
; PRIOR APPLICATION NUMBER: PCT/US99/09365
; PRIOR FILING DATE: 1999-04-29
; PRIOR APPLICATION NUMBER: 60/083,465
; PRIOR FILING DATE: 1998-04-29
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 9
; LENGTH: 2074
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-491-356C-9

Query Match 4.4%; Score 137.5; DB 4; Length 2074;
Best Local Similarity 18.0%; Pred. No. 0.001;
Matches 137; Conservative 118; Mismatches 226; Indels 281; Gaps 36;

QY 17 NPLQRHQSGORRELPRKIVETESTSMITTGSKQPO-----FLKSIDELAAFSVAVETP 72
Db 1215 DPLVSSAQAR-----LMQLICYPHRLDNDGPNQORIKRILKLDQW-----TM 1262
QY 73 KROFDLOKHIESIENAIKSLGNGVLAARNNFHQPMLSPRRNVSVETTVTSQPS 132
Db 1263 ROSSLELMQIKQTNT-----ENSLI-----ENIA-KATIEVFOQS 1299
QY 133 QIVPET-----SNKPEGGRMCELMCSKGLRKYIVANTSDQAKLMEETPSALKAKPAK 188
Db 1300 AETGSSSGSTASNPSSSKTKPVLSS---LER---SGVWLAPLIAKLTFSVQ-----GH 1348
QY 189 VLDCTGKFLQGR---AFTKESPMSSARQVSLILLESFL-----LMPDRGKGKVKIESW 240
Db 1349 VLKAGEELEKQHLGSGSRKERDRQKQKMSLLSQQFSLVTLCKGQDQREGILLAS 1408
QY 241 IKDEAETAAVAWR-----KSLMTE-----GGL-----AAAEKMDARGLL 275
```

```
Db 1409 LHSQVHOIWNRENQYLDCKPKQLMHEALKRLNLVGMFDTVQORSTQOTTEWAQLLL 1468
QY 276 LVACFGV---PSNFRSTDLILIRMSGNEIAGALKRSQFLVPMVSGIVESSIKRGMH 331
Db 1469 EIIISGVDMQSNELFTTVDMLSVLNGTLAADV-----SSISQG---1510
QY 332 EALEMVTYTFGMEKDFSAALVLTSL-----KMSKESPERAKRKAQSPFLAFKEAAT-----381
Db 1511 -----SMEENKRAYMNLVKLQKDLGERQSDSLEKVHQLLPLPKQNRDVITCEPQG 1561
QY 382 -----KQLAVLSSV-----MQMETHKLDPAKELPGWQIKQEIIVSLEK-----DTL 422
Db 1562 SLITKGNKLAGDISIPFKEGLOVSTQKLTSP-----WELFEGLPKSTAPLSWAWFGTV 1615
QY 423 OLDKEM---BEKARSLMBEAAALAKRMYNQOIK-RPR---LSPMEMPPVTSSSYPIY- 474
Db 1616 RVDRRVARGEQQRL-----LYHTHLRPRRAYYLEPLPLPPEDEPPAPALL 1664
QY 475 -----RDRSFPQORDDQDEISALVSSYLGPSSTSPHRSRSPSPVMDPLPHGGL-----522
Db 1665 EPEKKAPEPPKTDKPGAAPPSTBEERKKKSTGKKRSOPATKNEDYG-----MGPR 1715
QY 503 SFPHRSRRSPPEYM-----VPLPHGG-----522
Db 1716 SGPGVTVPPDLLHANPGSISHLISYQSSMGLYTQNPPLAGGPRVDYPRVRLPMQKL 1775
QY 523 -----LORSVYAYEHLAPNSYSPGHGRHLRQYSPSLVHQ---RHPLQYSPPIHQ 571
Db 1776 PTRPTYPGVLTMTMTVMGLEPSYKTS-----VYRQOQTPVQOQRLQOQSQOQMLQ 1831
QY 572 ---QQLP---YGIQVYVRHSPSEERY-----LGLSNQKSP 600
Db 1832 SSVHQTWPSSSYGLQTSQLSPSLQGYTSVSHVGLQOQTGP 1873

RESULT 3
US-09-491-356C-8
; Sequence 8, Application US/09491356C
; Patent No. 6566061
; GENERAL INFORMATION:
; APPLICANT: Philibert, Robert A.
; APPLICANT: Ginns, Edward I.
; TITLE OF INVENTION: IDENTIFICATION OF POLYMORPHISMS IN THE PCTG4 REGION OF XQ13
; FILE REFERENCE: 9465.6US11
; CURRENT APPLICATION NUMBER: US/09/491,356C
; CURRENT FILING DATE: 2000-01-26
; PRIOR APPLICATION NUMBER: PCT/US99/09365
; PRIOR FILING DATE: 1999-04-29
; PRIOR APPLICATION NUMBER: 60/083,465
; PRIOR FILING DATE: 1998-04-29
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8
; LENGTH: 2023
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-491-356C-8

Query Match 4.0%; Score 123.5; DB 4; Length 2023;
Best Local Similarity 18.0%; Pred. No. 0.025;
Matches 136; Conservative 120; Mismatches 225; Indels 273; Gaps 36;

QY 17 NPLQRHQSGORRELPRKIVETESTSMITTGSKQPOFLKSIDELAAFSVAVETPKQF 76
Db 1156 DPLVSSAQAR-----LMQLICYPHRLDNDGPNQORIKRILQ-----NLDQWTRQSS 1207
QY 77 DDLQKH-ESTENAIKSLGNGVLAARNNFHQPMLSPRRNVSVETTVTSQPSQI 135
Db 1208 LELQMLKQIPNNEMSLLE-----NIA-KATIEVFOQSAT 1243
QY 136 VPET-----SNKPEGGRMCELMCSKGLRKYIVANISQAKLMEETPSALKAKPAKFVLD 191
```

Db 1244 GSSSGTASNPSSKTKFVLS--LER--SGVNLVAPLIAKLPTSYQ-----GHVLK 1292
Qy 192 CIGKPYLOGRR--ATTKEPMSAPQVSLILLEGFLM-----PDRGKGVK--- 236
Db 1293 AAGELEKGOHLGSSRKERDQKQKMSLLSQPFSLVLTCLGQDEQREGLLTSYS 1382
Qy 237 -----IESWIKDP-----AETAAVAMRKLMTGGGL-----AAAFKMDARGILLVA 278
Db 1353 QVHQIVNNRFDQVLDCKPKQLMHEALKRLNLVGVGMPDVTORSTQQTTEWAMILLEII 1412
Qy 279 CFV-----PSNFRSTDLDIRMSGNEBAGALKSQFLVPMVSGIVSSIKRGHIEAL 334
Db 1413 ISGTVMQSNNEFLTFTVDMLSVLINGTLAADM-----SSISQG----- 1451
Qy 335 EMVYTFGMEKFGAALVLSFLKMSX-----SPERAKRKAQSPFLAKFAAT----- 381
Db 1452 -----SMEENKEAYMNLAK--KLQELGERQSDSLEKVRQLPLPKQTRDVTICEPOGS 1503
Qy 382 -----KQLAVLSV-----MQMETHKLDPAKELPGWQIKQIYSLEK-----DTLQL 424
Db 1504 LIDTGNKNIAGPDSIFKKEGLVSTKQKISP-----WDLFGLKPSAPLSWGNVFTVRV 1557
Qy 425 DKEM-----EEKARSLMEEAALAKRMVNOQIK-RPR-----LSPMEMPPVTSSSYSPIY--- 474
Db 1558 DRRVARGEEOQLL-----LYHTLRPRPRAYYLEPLPEDEEPATLLEP 1606
Qy 475 -----RDRSPSQRDDQDEISALVSSYLGPTSPF 504
Db 1607 EKAPEPKTKFGAAPPSTEERKKXKTKKRSQATKTEDYG-----MGFGRS 1657
Qy 505 PHRSRSPRYM-----VPLPHGG----- 522
Db 1658 PYGVTPDILHHPGSIHLYVROGSLGYTQNPQPLFAGPRVDPRVPLPMQKLP 1717
Qy 523 -----LGRSVAYEHLAPNSYSPGHRLHRQYSPSLVHQO--RHPLOYSPPIHQO-- 571
Db 1718 RPTVGVLPITMTGVLGMELEPSYKTS-----VYRQOPAVPQGOQLRQLOQSQGMLGQSS 1773
Qy 572 --QQLP---VGIVRHSSEERYGLSNQSP 600
Db 1774 VHOMTPSSYGLQTSQGYTPYVS-HVGLQOHTGP 1806

RESULT 4

US-09-308-375-2
; Sequence 2, Application US/09308375
; Patent No. 6300117
; GENERAL INFORMATION:
; APPLICANT: Genencor International, Inc.
; TITLE OF INVENTION: Proteases From Gram-Positive Organisms
; FILE REFERENCE: GC394-PCT
; CURRENT APPLICATION NUMBER: US/09/308,375
; CURRENT FILING DATE: 1999-05-14
; EARLIER APPLICATION NUMBER: EP9719636.4
; EARLIER FILING DATE: 1997-09-15
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 2285
; TYPE: PRT
; ORGANISM: Bacillus subtilis
US-09-308-375-2

Query Match 3.9%; Score 122.5; DB 4; Length 2285;
Best Local Similarity 19.3%; Pred. No. 0.038;
Matches 83; Conservative 70; Mismatches 151; Indels 125; Gaps 15;
Qy 159 KYIVANISD---QAKLMEIPSAKLAKPAKFLVDCIGKFKYQGRAPTKSPMSARQ 215
Db 526 KTIPIRIGNOSSIKALEIGISVKTAGGAKSADLISE-----VAGKNDTLSDAQK 578
Qy 216 VSLILLESFLMPDRGKGVKIESWIKDEAETA--VAWRKRL----- 256

Db 579 QNTSIGVAGIYQLSRFNAMNFSIAQNAAKTAANTSGSAWSEQQKYADSLQARVNKLQN 639
Qy 257 -NTEGGLAAAEKMDARGLLLVACFGVPSNFRSTDLDIRMSG--NEIAGALKRSQFL 313
Db 639 NTEFAIASDA-----FISDGLIEFTQAAGSLLNASTGVIKSVGFL 680
Qy 314 VMVSGI-----VSSIKRGHIEALEMYTFGMEKFGAALV----- 351
Db 681 PFLAAVSTATILLSKNTRTLLASSLILGTFRAMGQFTLATAGLEAGMTRAAVASVLKAL 740
Qy 352 -----LTSF--LKMSESFERAKR-KAQSPFLAPKEAATKQLA 385
Db 741 RGLLVSTLVGGFAALGHALSLESISSFAEAKKAKDDEQSOQTNVEAITTNKSDTKLIQ 800
Qy 386 VLSVNMQCMETHKLDPAKELPGWQIKQIV-----SLEKDTQLDKMEEE 430
Db 801 QYKELQKVKESRLTSDBEEQYLVTOQLAOTFPALVKGYDSQGNAILKTNKELEKAIEN 860
Qy 431 KARSLSL--MEEAALAKRMVNOQIKRPLSPMEMPPVTSSSYSPI--YRDRSFPSQ---R 483
Db 861 TKEYLAKKQETRDSAKTIFEDASKEIKSKDEL-----KQYKQIADYNDKGRPKWDLIA 915
Qy 484 DDQDEISA 492
Db 916 DDDDYKVA 924

RESULT 5

US-09-060-410-2
; Sequence 2, Application US/09060410
; Patent No. 6165461
; GENERAL INFORMATION:
; APPLICANT: Cobb, Melanie
; APPLICANT: Hutchinson, Michele
; APPLICANT: Chen, Zhu
; TITLE OF INVENTION: TAO PROTEIN KINASES AND METHODS OF USE
; TITLE OF INVENTION: THEREFOR
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/060,410
; FILING DATE: 14-APR-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 860098.421
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1001 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-060-410-2

Query Match 3.9%; Score 120.5; DB 3; Length 1001;
Best Local Similarity 19.8%; Pred. No. 0.015;


```

QY 489 ELSALVSSYLGPSTSTFFPHRSRRSPYVWPLPHGGLGHSVYAYEHLAPNSYSPOCHGHELLHR 544
DB 907 -----ASSWSHN-----PTGSG-----PWGHPM-- 926
QY 549 QYSPSLVHGQRHPLQYSP-----PIHQOQLPYGIQVYVRHSPSEERYL----GLSNQRSP 600
DB 927 ---GCTPOAWGHPMCGGPQPGWGHSGPMQGVPRGSSIGVENSQALRRRTASGGRTEQGMS 983
QY 601 RSNs 604
DB 984 RSTs 987

RESULT 7
US-08-685-576-1
; Sequence 1, Application US/08685576
; Patent No. 5906819
; GENERAL INFORMATION:
; APPLICANT: Kaibuchi, Kozo
; APPLICANT: Iwamatsu, Akihiro
; APPLICANT: Nakano, Takeshi
; APPLICANT: Ito, Masaaki
; APPLICANT: Takahashi, No. 5906819uaki
; TITLE OF INVENTION: RHO TARGET PROTEIN RHO-KINASE
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/685,576
; FILING DATE: 24-JUL-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 7-325129
; FILING DATE: 20-NOV-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 8-17150
; FILING DATE: 05-JAN-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 8-131206
; FILING DATE: 26-APR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Bent, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 16887/843
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1388 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-685-576-1

Query Match 3.8%; Score 119.5; DB 2; Length 1388;
Best Local Similarity 20.3%; Pred. No. 0.033;
Matches 113; Conservative 99; Mismatches 209; Indels 137; Gaps 255
QY 17 NPLQRHQSQRREPLK---IVETES-TSMDDITGQ---SKQOFLKSID-ELAAFSVA 68
DB 549 NSGISTEKVNLQRODFTNAAARLRTQAESKQIQLESNNRDLQKNCI 608

```

QY	69	VETFK-----RQPDLOKHIES-----TENAIOSKLE-----SNGVVLAAARNNF	108
DB	609	LETAKLKEKEFNLQSGESERRDTHGSEIINDLQGRISGLEEDVNGKILLAKVELE	668
QY	109	HQPM-----LSPPRNNVYET--VTVSQPSQEIPTPSNKNPEGRWCMLSCGRLKY	160
DB	669	KRQLOERFTDLEKKNMEIDMTYQLKVQOSLE--QEETEHKATKARLAD-----KNK	720
QY	161	IYANISD--QAKLMEEIPSALK-----LAKSPAKFVLDICGKLYLQGRRAFTK	206
DB	721	IYESIEBAKSEAMKEMKKUSEERTUKQVKNLLEAEKRCISILDCDLK-----	769
QY	207	ESPMSSARQVSLTILESFLMLPPGRGKVKVIESWIKDEAETAANAWRKMLMTEGGLAAAE	266
DB	770	----QSQKINELLKQDVNLNEDVRNLTKIEQ-----ET-----QKRCLTQNDLKMQT	814
QY	267	----KWDARGLLLVACGCVSPNFRSTDLIDLTMSGSNEIAGALKRSQFLVPMVSGI	320
DB	815	QOVNTRKSEKQL-----KOENHLLLEM--KMSLEKQNAELRKEQDQDQOMKEL	862
QY	321	VESSTKRGMHTEALEMYVTFCGMDKFTSAALVLTSLFKMSKESF--ERAKRKAQSPKAFKE	378
DB	863	--QDQLEAEQVFTLYKTQVRELKEECEBTKLCKELQKKQLQDQERDSLAAQLEITLTK	921
QY	379	AATKQLAVLSVMQCNETHKLDPAKELPGWQIKEQ-----IYSLEK-----	419
DB	922	ADSEGLARSTAEQVYSLEKEKIMKEL--EIKEMMARHKQBELTEKDATIASLEETNRTL	978
QY	420	-----DTQLQDKEMESKARSLMSLMBEAAALKRMVYNOQIKRPLSPMEMPPVTSSSY	470
DB	979	TSDVANLANEKEELNNKLEKQAQOLSELKDEEISAAAIKAQFEKOLLTERTLTKTOAVNKL	1038
QY	471	SPIYRDRSPSQRRDDQD 488	
DB	1039	ABE--MWRKEPVKRGNDTD 1055	
RESULT 8			
US-08-261-304-2			
; Sequence 2, Application US/08261304			
; Patent No. 5708147			
; GENERAL INFORMATION:			
; APPLICANT: Cybulecky, Myron I.			
; APPLICANT: Gimbirne, Michael A.			
; APPLICANT: Collins, Tucker			
; TITLE OF INVENTION: Mononuclear Leukocyte Directed			
; TITLE OF INVENTION: Endothelial Adhesion Molecule Associated with			
; TITLE OF INVENTION: Atherosclerosis			
; NUMBER OF SEQUENCES: 13			
; CORRESPONDENCE ADDRESS:			
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox			
; STREET: 1225 Connecticut Avenue, N.W.			
; CITY: Washington			
; STATE: District of Columbia			
; COUNTRY: United States of America			
; ZIP: 20036			
; COMPUTER READABLE FORM:			
; MEDIUM TYPE: Floppy Disk			
; COMPUTER: IBM PC compatible			
; OPERATING SYSTEM: PC-DOS/MS-DOS			
; SOFTWARE: Ascii			
; CURRENT APPLICATION DATA:			
; APPLICATION NUMBER: US/08/261,304			
; FILING DATE:			
; CLASSIFICATION: 435			
; PRIOR APPLICATION DATA:			
; APPLICATION NUMBER: 07/649,565			
; FILING DATE: 01-FEB-1991			
; APPLICATION NUMBER: U.S. 07/487,038			
; FILING DATE: 02-MAR-1990			
; ATTORNEY/AGENT INFORMATION:			

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/533.306A
FILING DATE: September 25, 1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Smith, Dean F.
REGISTRATION NUMBER: 36683
REFERENCE/DOCKET NUMBER: 2115-00869COB
TELECOMMUNICATION INFORMATION:
TELEPHONE: (810) 641-1600
TELEFAX: (810) 641-0270
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 816 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-533-306A-6

Query Match	3.7%;	Score 113.5;	DB 2;	Length 816;
Best Local Similarity	21.1%;	Pred. No. 0.052;		
Matches 100: Conservative	76;	Mismatches 190;	Indels 109;	Gaps 20;

Qy	24	QSEORRELPIVETES-----TSMDDITIGSQPOPLKSIDELAFAFSAVATEFK	73
Ddb	218	QLEBENSLODDEMEAKONLERHISTINQLSDSK-----KKIQDPATVTAEHGK	272
Qy	74	RQFDLQKHTESTENAIOSKLNSGVVLAAARNNHPQMSPPRNNVSIVETTIVSOPSQ	133
Ddb	273	KRF---QKEIENLTQQYEKAAYDKLETKN-----RLQCELDLVVDLNDQR	316
Qy	134	EIVPETSNGEGGEMELCMSCGLRYKIYANISD--OAKLMEEIPTSALKIAK--EPAPKV	189
Ddb	319	OLVSNLEKKOR--KFTQLLAEEKNISCKYADERDRAEAAREKETKALSLARAEAA---	373
Qy	190	LDCIGKFYLGRRAFTKESPMSSARQVLSLLIESFLMLPDRCGKGVKIESWIKDEAETA	249
Ddb	374	-----LEAKELBRTNKMLKAEMEDIVSSKD-----DYGNVHELEK-SKGALETQM	419
Qy	250	VAMRKRL-MTEGGILAAAEKMDAGLLLLVACFGVPSNFIRSTDLLDIIMSGSGNEIAGALK	308
Ddb	420	EEMKTQLEELEDDELQASE--DAK-LRLEVNMQAALQGPER----DLQARDQONE-----EK	468
Qy	309	RSQFLVPWVGSIVESTIKGMHTALEMUYTTGEMEDKESAAILVLSFLMKSKESFERAKR	368
Ddb	469	RRQ-----LQROQLH-----BYETELEDERNERALAAAANKKULEGDLKULEL	509

```

RESULT 10
US-08-742-923A-6
; Sequence 6, Application US/08742923A
; Patent No. 5869611
; GENERAL INFORMATION:
; APPLICANT: Liu, Pu
; APPLICANT: Collins, Francis S.
; APPLICANT: Siciliano, Michael J.
; APPLICANT: Claxton, David
; TITLE OF INVENTION: Markers for Detection of Chromosome 16
; TITLE OF INVENTION: Rearrangements
; NUMBER OF SEQUENCES: 14
;

```

```

RESULT 9
US-08-533-306A-6
; Sequence 6, Application US/08533306A
; Patent No. 5837457
;
; GENERAL INFORMATION:
;
; APPLICANT: Liu, Pu
; APPLICANT: Collins, Francis S.
; APPLICANT: Siciliano, Michael J.
; APPLICANT: Claxton, David
;
; TITLE OF INVENTION: Markers for Detection of Chromosome 16
; TITLE OF INVENTION: Rearrangements
;
; NUMBER OF SEQUENCES: 14
;
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Harness, Dickey & Pierce, P.L.C.
; STREET: P.O. Box 828
; CITY: Bloomfield Hills
; STATE: MI
; COUNTRY: USA
; ZIP: 48303

```

; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Harness, Dickey & Pierce, P.L.C.
 ; STREET: P.O. Box 828
 ; CITY: Bloomfield Hills
 ; STATE: MI
 ; COUNTRY: USA
 ; ZIP: 48303
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/742,923A
 ; FILING DATE: No. 5969611ember 1, 1996
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Smith, DeAnn F.
 ; REGISTRATION NUMBER: 36683
 ; REFERENCE/DOCKET NUMBER: 2115-00869DVC
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (810) 641-1600
 ; TELEFAX: (810) 641-0270
 ; INFORMATION FOR SEQ ID NO: 6:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 816 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-742-923A-6

Query Match 3.7%; Score 113.5; DB 2; Length 816;
 Best Local Similarity 21.1%; Pred. No. 0.052;
 Matches 100; Conservative 76; Mismatches 190; Indels 109; Gaps 20;

Qy	24	QSEQRRELPIKIVETES-----TSMDTIGSKQPOFLKSIDELAASFVAVERFK 73
Db	218	QLEHERNSLOQDDEMEAKQNLERHISTLNQLSDSK-----KKLQDFASVTEALEEGK 272
Qy	74	RQFDDLOKHIESIENAIKSLKESNGVVLAAARNNNHQPMLSPRRNVSVETTVTSQPSQ 133
Db	273	KRF---QKEIENLTQOYEEKAAAYDKLETKN-----RLOQELDDLVVDLDNQR 318
Qy	134	EIVPETSNGPEGRMCELMCSKGLRKVIYANISD--QAKLMBEIPSAKLAK--BPAKFV 189
Db	319	QLVSNLEKKQR--KFDOLLAEEKNISSKYADERDRAEAAREKETKALSARALBEA--- 373
Qy	190	LDICIGFYLGQRAFTKESPMSSARQVSLILLESFLMPDRGKGVKIESWIKDEAETAA 249
Db	374	-----LEAKEBELRTNKLKAEEDLVSSKD-----DVGKNVHELEK-SKRALETQM 419
Qy	250	VAMRKRL-MTEGGILAAEKMDARGLLLVACFGVPSNFRSTDLDIRMSGNETAGALK 308
Db	420	EEMTKOLEELEDLOASE--DAK-LRLEVNQALKQOGER-----DLQARDEQNE-----EK 468
Qy	309	RSQFLVPMVSGIVESSIKRGMHIEALEMVYTFGMEDKFSAAVLTSFLKMSKESPERAKR 368
Db	469	RRQ-----LQRLH-----EYETELEDERNERALLAAAKKLEGLDKLDEL 509
Qy	369	KAQSLAPFKEAATKQLAVLSSVWQ-----CMETHKLDPAKELPGWQ 409
Db	510	QADSAIKGREEAIKQLRKLOAKQDKDFORELEDARASRDEIFATAKENEKAKSLEADLMQ 569
Qy	410	IKEQIVSLEKDTLQDKEMEKEKARSLSLMEEAALAKRMVNNQIKR---PRLSPME 461
Db	570	LQEDLAAARARKQADLEKEELAEELA-----SSLSGRNALQDEKRLERLARIAQLE 620

RESULT 11
 US-08-533-306A-4
 ; Sequence 4, Application US/08533306A
 ; Patent No. 5837457
 ; GENERAL INFORMATION:

; APPLICANT: Liu, Pu
 ; APPLICANT: Collins, Francis S.
 ; APPLICANT: Siciliano, Michael J.
 ; APPLICANT: Claxton, David
 ; TITLE OF INVENTION: Markers for Detection of Chromosome 16
 ; TITLE OF INVENTION: Rearrangements
 ; NUMBER OF SEQUENCES: 14
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Harness, Dickey & Pierce, P.L.C.
 ; STREET: P.O. Box 828
 ; CITY: Bloomfield Hills
 ; STATE: MI
 ; COUNTRY: USA
 ; ZIP: 48303
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/533,306A
 ; FILING DATE: September 25, 1995
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Smith, DeAnn F.
 ; REGISTRATION NUMBER: 36683
 ; REFERENCE/DOCKET NUMBER: 2115-00869COB
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (810) 641-1600
 ; TELEFAX: (810) 641-0270
 ; INFORMATION FOR SEQ ID NO: 4:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 885 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-533-306A-4

Query Match 3.7%; Score 113.5; DB 2; Length 885;
 Best Local Similarity 21.1%; Pred. No. 0.06;
 Matches 100; Conservative 76; Mismatches 190; Indels 109; Gaps 20;

Qy	24	QSEQRRELPIKIVETES-----TSMDTIGSKQPOFLKSIDELAASFVAVERFK 73
Db	287	QLEHERNSLOQDDEMEAKQNLERHISTLNQLSDSK-----KKLQDFASVTEALEEGK 341
Qy	74	RQFDDLOKHIESIENAIKSLKESNGVVLAAARNNNHQPMLSPRRNVSVETTVTSQPSQ 133
Db	342	KRF---QKEIENLTQOYEEKAAAYDKLETKN-----RLOQELDDLVVDLDNQR 387
Qy	134	EIVPETSNGPEGRMCELMCSKGLRKVIYANISD--QAKLMBEIPSAKLAK--BPAKFV 189
Db	388	QLVSNLEKKQR--KFDOLLAEEKNISSKYADERDRAEAAREKETKALSARALBEA--- 442
Qy	190	LDICIGFYLGQRAFTKESPMSSARQVSLILLESFLMPDRGKGVKIESWIKDEAETAA 249
Db	443	-----LEAKEBELRTNKLKAEEDLVSSKD-----DVGKNVHELEK-SKRALETQM 488
Qy	250	VAMRKRL-MTEGGILAAEKMDARGLLLVACFGVPSNFRSTDLDIRMSGNETAGALK 308
Db	489	EEMTKOLEELEDLOASE--DAK-LRLEVNQALKQOGER-----DLQARDEQNE-----EK 537
Qy	309	RSQFLVPMVSGIVESSIKRGMHIEALEMVYTFGMEDKFSAAVLTSFLKMSKESPERAKR 368
Db	538	RRQ-----LQRLH-----EYETELEDERNERALLAAAKKLEGLDKLDEL 578
Qy	369	KAQSLAPFKEAATKQLAVLSSVWQ-----CMETHKLDPAKELPGWQ 409
Db	579	QADSAIKGREEAIKQLRKLOAKQDKDFORELEDARASRDEIFATAKENEKAKSLEADLMQ 638
Qy	410	IKEQIVSLEKDTLQDKEMEKEKARSLSLMEEAALAKRMVNNQIKR---PRLSPME 461
Db	639	LQEDLAAARARKQADLEKEELAEELA-----SSLSGRNALQDEKRLERLARIAQLE 689

RESULT 12
US-08-742-923A-4
; Sequence 4, Application US/08742923A
; Patent No. 5869611
; GENERAL INFORMATION:
; APPLICANT: Liu, Fu
; APPLICANT: Collins, Francis S.
; APPLICANT: Siciliano, Michael J.
; APPLICANT: Claxton, David
; TITLE OF INVENTION: Markers for Detection of Chromosome 16
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Harness, Dickey & Pierce, P.L.C.
; STREET: P.O. Box 828
; CITY: Bloomfield Hills
; STATE: MI
; COUNTRY: USA
; ZIP: 48303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/742,923A
; FILING DATE: No. 5869611ember 1, 1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Smith, Deann F.
; REGISTRATION NUMBER: 36683
; REFERENCE/DOCKET NUMBER: 2115-00869DVC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (810) 641-1600
; TELEFAX: (810) 641-0270
; INFORMATION FOR SEQ ID NO: 4:
; LENGTH: 885 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-742-923A-4
Query Match 3.7%; Score 113.5; DB 2; Length 885;
Best Local Similarity 21.1%; Pred. No. 0.06;
Matches 100; Conservative 76; Mismatches 190; Indels 109; Gaps 20;
QY 24 QGEORRELPKIVETES-----TSMDITIGSKQPOFLKSIDELAAFSAVAVETPK 73
Db 287 QLEERNSLQDQIDEMEAKQNLERHISTLNQISDSK-----KKIQDFSTVEALEGK 341
QY 74 RQFDDLQKHIESIENAIKSLKESNGVLAARNNNPHQPMSPNNVSVETTVTSQPSQ 133
Db 342 KRF--QKEIENLTQYEEKAAAYDKLETKN-----RLQELDDLVDLDNQR 387
QY 134 EIVPETSNGPEGRWCMLSCSKGLRKYIYANISD--QAKLMEETPSALKAK--EPAKFV 199
Db 388 QLVSNLEKQK--KFDQLAEKNTSSKYADRDRAEAEAEKTKALSARALEEA---442
QY 190 LDCIGKFIYLGRAFTKESPMSSARQVSLILLESFLLMPDRGKGKVKIETSWIDEAETA 249
Db 443 -----LEAKELEERTNKMKAEMEDLVSSKD-----DVGKNVHELEK-SKRALETQM 488
QY 250 VAWRKEL-MTEGGLAAAEKQWDARGILLVACFGVPNSFRSTDLLDLMSSGNEIAGALK 308
Db 489 ENKKTQLELEDELQASE--DAK-LRLEVNQALKGQFER-----DQARDEQNE-----EK 537
QY 309 RSQFLVPMVSGIVGESIKGMHIEALEMVTYFGMEDKFSAAVLTSFLKMSKESFEAKR 368
Db 538 RQK-----LQRLH-----EYTELEDERNERALAAAKKLEGLDKLEL 578

QY 369 KAQSLAPKFAATYQTLAVLSSVMQ-----CMETHKLDPAKELPGHQ 409
Db 579 QADSAIKGREAEIXQLRKLQAKMDFQRELEDPARASDEIFAFAKENEKAKSLEADLMQ 638
QY 410 IKEOIVSLEKDTLQDKEMBEKARSLSIMEEAAALAKRMVNNQIKR---PRLSPME 461
Db 639 LQEDLAAARAKQADLEKELEBELA-----SSUSGSENALQDEKRELEARIAGLE 689
RESULT 13
US-08-353-700-1
; Sequence 1, Application US/08353700
; Patent No. 5599919
; GENERAL INFORMATION:
; APPLICANT: YEN, TIMOTHY J.
; APPLICANT: RATNER, JEROME B.
; TITLE OF INVENTION: NUCLEIC ACID ENCODING A
; TITLE OF INVENTION: TRANSIENTLY-EXPRESSED KINETOCHORE PROTEIN,
; TITLE OF INVENTION: AND METHODS OF USE
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DANN, DORFMAN, HERRELL AND SKILLMAN
; STREET: 1601 MARKET STREET, SUITE 720
; CITY: PHILADELPHIA
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2307
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/353,700
; FILING DATE: 09-DEC-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: REED, JANET E.
; REGISTRATION NUMBER: 36,252
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 563-4100
; TELEFAX: (215) 563-4044
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3248 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: HUMAN
US-08-353-700-1
Query Match 3.6%; Score 112; DB 1; Length 3248;
Best Local Similarity 19.5%; Pred. No. 0.76;
Matches 115; Conservative 105; Mismatches 219; Indels 150; Gaps 24;
QY 11 QPTTANPLLQHQSEORRR-----ELPKIVETESTMDITIGSKQPOFLKSIDE- 61
Db 2530 EKSSTAMEMLQTLKELNERYAALHNDQEAACKAKE-QNLSQVECLELEKALQLOGIDEA 2588
QY 62 -----LAAFSVAVETPKRFQDLDQKHIESIENAI-----SKL----- 94
Db 2589 KNNYIVLOSSVKGILIOEVEDQKQKLEKDEISLKNQIQDQQLVSKLSQVEGEHLWK 2648
QY 95 ESNQVVLAAARN--NNFHQPMSPNNVSVETTVTSQPSQEIIVPETSNGPEGRWCML 152
Db 2649 EQN---LELRNLTVELEKQIVLQSKNASIQDTLEVLQSSYK---NLENELELTCKDKMS 2702
QY 153 CSKGLRKYIYANISDQAKLMEETPSALKAKAEKAFVLDICIG-KFYLGERRAFTKESPM 211

Db 2703 FVEKVNMM-----TAKETELQREMH-----EMAQKTAELOEELSSEKRNLAGELQELLEIKS 2755
Qy 212 SARQVSLILLESFLL---MPDRGKGKVKIESWIKDEA----- 245
Db 2756 SKDQLKELTLENSELKSLDCMVKDQVEKSGKVRBEIAEYQLRLHEAEKQKQALLDITNK 2815
Qy 246 --ETAAVAKRLMTGGLAAAKMDARGILLLVACFGVPSFRSTDLLDIRSGSNEI 303
Db 2816 QYVEIQTVREKLTSEKLESSLQKLE-----IDLLK-SKKEEL 2852
Qy 304 AGALKRSQFLVPMVSGIVESSIKRGHIEBALENVYTFGMEDKFSAA---LVLTSFLKMSK 360
Db 2853 NNSLKAT-----TQILEELKXTKM--DNLKYNQLKKENERAQQKMKLLIKSCQKLE 2903
Qy 361 ESFERAKRKAQSLAPKAEATKQLAVLSSVMQCMETHKLDPAK--ELPWQIKQEIIVSLE 418
Db 2904 E-----KEILQKELSQAQAEKQKGTGTMTKVDLT-TEIKELKETLE 2947
Qy 419 KDTLQDKEMEKARSLSLMEEAALAKMYNQOI-----KPRLSPMEMPVPTSSS 469
Db 2948 EKTKEADYLDKYCSLLISHEKLEKAKEMLETOVAHLCSQQSKQDSRGSPLLGPVPGPS 3007
Qy 470 YSPIYDRSPFS---QRDDQDDEISALVSSYLGPS-----TSFPHRSRS 511
Db 3008 PIPSVTEKRLSSGQNKASGRQRSSGIWENGRGPTPATPESFSKSKKA 3056
RESULT 14
PCT-US95-16216-1
; Sequence 1, Application PC/TUS9516216
; GENERAL INFORMATION:
; APPLICANT: Yen, Timothy J.
; APPLICANT: Rattner, Jerome B.
; TITLE OF INVENTION: Nucleic Acid Encoding a Transiently
; TITLE OF INVENTION: Expressed Kinetochores Protein, and Methods of Use
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pann, Dorfman, Herrell and Skillman
; STREET: 1601 Market Street Suite 720
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2307
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/16216
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/353,700
; FILING DATE: 09-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Reed, Janet E.
; REGISTRATION NUMBER: 36,252
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 563-4100
; TELEFAX: (215) 563-4044
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3248 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; HYPOTHEetical: NO
; ANTI-SENSE: NO
PCT-US95-16216-1

Query Match 3.6%; Score 112; DB 5; Length 3248;

Best Local Similarity 19.5%; Pred. No. 0.76;
Matches 115; Conservative 105; Mismatches 219; Indels 150; Gaps 24;
Qy 11 QPTTANPLQLQSHQSEQR---ELPKIVETESTMDITIGOSKQPFLLKSID- 61
Db 2530 EKSSSTAMEMLTQKELNERNVAALHNDQEAACKAE-QNLSQVECLEKAKQLQLGLDEA 2588
Qy 62 -----LAASFVAVETPKRQDDDLQKHIESIENAI-----SKL----- 94
Db 2589 KNNYIVLOSVMKGLQOEVDGKQLEKDEISRLKNQIQOEQLVSKLSQVEGHEQLNK 2648
Qy 95 ESGGVVLAARN--NNFHOPMLSPPPNNVSVTTVTSQPSQEIIVPETSNNKPEGGRMCELM 152
Db 2649 EQN---LELRNLTVELEQKIQVLOSKNASLQDTLEVLSQSYK--NLNLELELTCKMDKMS 2702
Qy 153 CSKGLKYYIYANISQAKLMBEIPALAKLAKPAKFLVDCIG-KFYLCGRRAFTKESFMS 211
Db 2703 FVEKVNMM-----TAKETELQREMH-----EMAQKTAELOEELSSEKRNLAGELQELLEIKS 2755
Qy 212 SARQVSLILLESFLL---MPDRGKGKVKIESWIKDEA----- 245
Db 2756 SKDQLKELTLENSELKSLDCMVKDQVEKSGKVRBEIAEYQLRLHEAEKQKQALLDITNK 2815
Qy 246 --ETAAVAKRLMTGGLAAAKMDARGILLLVACFGVPSFRSTDLLDIRSGSNEI 303
Db 2816 QYVEIQTVREKLTSEKLESSLQKLE-----IDLLK-SKKEEL 2852
Qy 304 AGALKRSQFLVPMVSGIVESSIKRGHIEBALENVYTFGMEDKFSAA---LVLTSFLKMSK 360
Db 2853 NNSLKAT-----TQILEELKXTKM--DNLKYNQLKKENERAQQKMKLLIKSCQKLE 2903
Qy 361 ESFERAKRKAQSLAPKAEATKQLAVLSSVMQCMETHKLDPAK--ELPWQIKQEIIVSLE 418
Db 2904 E-----KEILQKELSQAQAEKQKGTGTMTKVDLT-TEIKELKETLE 2947
Qy 419 KDTLQDKEMEKARSLSLMEEAALAKMYNQOI-----KPRLSPMEMPVPTSSS 469
Db 2948 EKTKEADYLDKYCSLLISHEKLEKAKEMLETOVAHLCSQQSKQDSRGSPLLGPVPGPS 3007
Qy 470 YSPIYDRSPFS---QRDDQDDEISALVSSYLGPS-----TSFPHRSRS 511
Db 3008 PIPSVTEKRLSSGQNKASGRQRSSGIWENGRGPTPATPESFSKSKKA 3056
RESULT 15
US-08-328-254-6
; Sequence 6, Application US/08328254
; Patent No. 5710022
; GENERAL INFORMATION:
; APPLICANT: Zhu, Xueliang
; APPLICANT: Lee, Wen-Hwa
; TITLE OF INVENTION: A No. 5710022el Nuclear Mitotic Phosphoprotein
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/328,254
; FILING DATE: 24-OCT-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/141,239
; FILING DATE: 22-OCT-1993
; ATTORNEY/AGENT INFORMATION:

NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-CJ 1191
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 2482 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-328-254-6

Query Match 3.5%; Score 110; DB 1; Length 2482;
Best Local Similarity 18.9%; Pred. No. 0.76;
Matches 140; Conservative 120; Mismatches 272; Indels 208; Gaps 30;
QY 11 QPTTANPLLQRHQEQRR-----ELPKIVETSTMDITIGSQKQPOFLKSID- 61
DB 1802 EKSSTAMEMLQTLKELNVAALHNDQEAACKAE-QNLSSQVECLELEKALQLGLDEA 1860
QY 62 -----LAAPSVAVETFKRPFDDLQKHIESIENAIID-----SKL----- 94
DB 1861 KNYIVLOSSVNGLTQVEDGQKLEKDEISLKNQIQDQEQLVSKLSQVEGEHQLMK 1920
QY 95 ESGVVLAAARN--NNFHQPLMSPRNNVSVETTVTSQPSQEIIVFETSNKPEGGRMCELM 152
DB 1921 EQN---LELRNLTVLEQKIQVLSQKNASLQDTLEVLQSSYK---NLNELELTQMDKMS 1974
QY 153 CSKGLRKIYANISDQAKLMEIIPSAKLAKPAKFVLDICG-KFYLGQRRATFKESPM 211
DB 1975 FVEKNKMM---TAKETELQREMH---EMAQTAELQBELSGEKRLAGELQLLLEIKS 2027
QY 212 SARQVSLTILESFL---MPDRGKGKVKIESWIKDEA----- 245
DB 2028 SKDQLKELTLENSEKSLDCMHWQDQVEKGEKVEETAEYQRLRHEAEKKHQALLDNTK 2087
QY 246 --ETAAVAKRELMTGGLAAXKMDARGLLLVACFGVPSNPFSTDLDDLIRMSGNEI 303
DB 2088 QYEVETIQTREKLSKESCLSSQKLE-----IDLK-SKEEL 2124
QY 304 AGALKRSQFLVPMVSGIVESSIKRGHIEALEMVTTFGMEKFSAA---LVLTSLFKMSK 360
DB 2125 NNSLKAT-----TQILEELKTKV--DNLKYVNLKKNENERAQGKMKLLIKCKQLEE 2175
QY 361 ESFERAKRAQSPAPAKAAIKQALVLSVMQCMETHKLDPAK--ELPGWIKQIVSL 418
DB 2176 E-----KEILOKLSQQAQKOKTGTVMDTKYDELT-TEIKELKETLE 2219
QY 419 KDTQLDKEMEEKARSLSLMEEAALAKEMVYNOQI-----KRPRLSPMEMPPVTSSS 469
DB 2220 EKTKEADEVLKYCSLLISHEKLEKEMLETVAHLCSSQKQSDSRGSPILGVPVPGPS 2279
QY 470 YSPIVDRSFFS---QRDDQDEISALVSSYLGPS-----TSFPHRSRRS----- 511
DB 2280 PIPSVTERKLSSGQNKASGRQSSGIWENGSGTPTATPESFKSKSKAVMSGIHPAEDT 2339
QY 512 -----PEVM---VPLPHGG-----LGRSVYA-----YEH 532
DB 2340 EGTEFEPEGLPEVVKKGADITGKTSPYILRTTMAITSPRLAAQKALSLPLSGKEN 2399
QY 533 LAPNSYSPGHRLHRQVSPSLVHQCRLPQVSPPIHQCQQLPYGIORVYVHSPSEERYL 592
DB 2400 LAESSKPTAGGSR-----SQVKVAQRSPVDSGTTILREFTTKSVFVNNLPERSPTDSPRE 2454
QY 593 GLSNQRS---PRNSGLDEK 609
DB 2455 GURVKRGRIVSPFKAGLESK 2474

Search completed: August 14, 2003, 10:41:27
Job time : 33 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 14, 2003, 10:40:05 ; Search time 59 seconds
(without alignments)
1352.220 Million cell updates/sec

Title: US-09-890-475-1

Perfect score: 3104

Sequence: 1 MSNPPTVAQAQTTTANPLL.....RYLGLSNQSPRNSSLDPX 609

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 492763 seqs, 131003257 residues

Total number of hits satisfying chosen parameters: 492763

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Published Applications AA.*
- 1: /cgn2_6/prodata/1/pubpaa/US07_PUBCOMB.pep.*
 - 2: /cgn2_6/prodata/1/pubpaa/PCT_NEW_PUB.pep.*
 - 3: /cgn2_6/prodata/1/pubpaa/US06_PUBCOMB.pep.*
 - 4: /cgn2_6/prodata/1/pubpaa/US06_PUBCOMB.pep.*
 - 5: /cgn2_6/prodata/1/pubpaa/US07_NEW_PUB.pep.*
 - 6: /cgn2_6/prodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
 - 7: /cgn2_6/prodata/1/pubpaa/US08_NEW_PUB.pep.*
 - 8: /cgn2_6/prodata/1/pubpaa/US08_PUBCOMB.pep.*
 - 9: /cgn2_6/prodata/1/pubpaa/US09_PUBCOMB.pep.*
 - 10: /cgn2_6/prodata/1/pubpaa/US09_PUBCOMB.pep.*
 - 11: /cgn2_6/prodata/1/pubpaa/US09_PUBCOMB.pep.*
 - 12: /cgn2_6/prodata/1/pubpaa/US10A_PUBCOMB.pep.*
 - 13: /cgn2_6/prodata/1/pubpaa/US10A_PUBCOMB.pep.*
 - 14: /cgn2_6/prodata/1/pubpaa/US10B_PUBCOMB.pep.*
 - 15: /cgn2_6/prodata/1/pubpaa/US10C_PUBCOMB.pep.*
 - 16: /cgn2_6/prodata/1/pubpaa/US10_NEW_PUB.pep.*
 - 17: /cgn2_6/prodata/1/pubpaa/US60_NEW_PUB.pep.*
 - 18: /cgn2_6/prodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	144	4.6	725	11	US-09-978-309A-47
2	132.5	4.3	888	11	US-09-893-519A-73
3	129	4.2	771	15	US-10-204-887-110
4	127.5	4.1	786	11	US-09-291-417-23
5	127.5	4.1	1001	11	US-09-291-417-31
6	127.5	4.1	1005	15	US-10-177-293-254
7	127.5	4.1	3256	10	US-09-919-172-98
8	127.5	4.1	3256	11	US-09-919-039-21
9	124.5	4.0	1203	15	US-10-097-340-43
10	124	4.0	426	15	US-10-156-761-12795
11	122.5	3.9	2285	10	US-09-332-183A-2
12	115.5	3.7	1001	15	US-10-128-714-3240
13	114.5	3.7	862	15	US-10-177-293-391
14	114.5	3.7	1938	15	US-10-171-311-164
15	114.5	3.7	1945	11	US-09-927-597-2

16	114.5	3.7	1972	15	US-10-171-311-162	Sequence 162, App
17	114.5	3.7	1979	11	US-09-927-597-4	Sequence 4, Appl
18	113	3.6	647	9	US-09-841-132-436	Sequence 436, App
19	112.5	3.6	892	12	US-10-205-219-102	Sequence 102, App
20	112.5	3.6	914	15	US-10-128-714-8595	Sequence 8595, Ap
21	111.5	3.6	1069	15	US-10-146-473-77	Sequence 77, Appl
22	111.5	3.6	26926	10	US-09-759-508B-2	Sequence 2, Appl
23	111.5	3.6	2139	9	US-09-727-384-6	Sequence 6, Appl
24	110.5	3.6	2139	15	US-10-023-219-4	Sequence 4, Appl
25	110	3.5	1101	15	US-10-177-293-106	Sequence 106, App
26	109.5	3.5	660	14	US-10-007-693-139	Sequence 139, App
27	109	3.5	578	9	US-09-159-469-50	Sequence 50, Appl
28	109	3.5	578	9	US-09-798-042-50	Sequence 50, Appl
29	108.5	3.5	748	9	US-09-864-761-43244	Sequence 43244, A
30	108	3.5	664	9	US-09-823-038A-48	Sequence 48, Appl
31	108	3.5	1867	11	US-09-824-574-5	Sequence 5, Appl
32	107.5	3.5	749	10	US-09-833-205-4	Sequence 4, Appl
33	107.5	3.5	1233	11	US-09-291-417-89	Sequence 89, Appl
34	106.5	3.4	489	9	US-09-876-889-350	Sequence 350, App
35	106	3.4	1618	10	US-09-963-875-1	Sequence 1, Appl
36	106	3.4	1618	15	US-10-136-891-2	Sequence 2, Appl
37	106	3.4	1618	15	US-10-120-887-1	Sequence 1, Appl
38	105.5	3.4	945	9	US-09-745-763-191	Sequence 191, App
39	105	3.4	594	12	US-10-137-870-10	Sequence 10, Appl
40	105	3.4	594	12	US-10-140-018-10	Sequence 10, Appl
41	105	3.4	594	12	US-10-140-021-10	Sequence 10, Appl
42	105	3.4	594	12	US-10-140-274-10	Sequence 10, Appl
43	105	3.4	594	12	US-10-140-471-10	Sequence 10, Appl
44	105	3.4	594	12	US-10-140-807-10	Sequence 10, Appl
45	105	3.4	594	12	US-10-140-922-10	Sequence 10, Appl

ALIGNMENTS

RESULT 1

US-09-978-309A-47
; Sequence 47, Application US/09978309A
; Publication No. US20030100490A1
; GENERAL INFORMATION:
; APPLICANT: Cruz, Tony
; APPLICANT: Pastreky, Aleksandra
; APPLICANT: Turley, Eva A.
; TITLE OF INVENTION: Compositions and Methods for Treating Cellular Response to Injury and Other Proliferating Cell Disorders Regulated by Hyaladherin and Hyaluronans
; TITLE OF INVENTION: Hyaladherin and Hyaluronans
; FILE REFERENCE: 033352-010
; CURRENT APPLICATION NUMBER: US/09/978,309A
; CURRENT FILING DATE: 2002-04-04
; PRIOR APPLICATION NUMBER: US 09/685,010
; PRIOR FILING DATE: 2000-10-05
; PRIOR APPLICATION NUMBER: US 09/541,522
; PRIOR FILING DATE: 2000-04-03
; PRIOR APPLICATION NUMBER: US 60/127,457
; PRIOR FILING DATE: 1999-04-01
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 47
; LENGTH: 725
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-978-309A-47

Query Match 4.6%; Score 144; DB 11; Length 725;
Best Local Similarity 19.6%; Pred. No. 0.00094;
Matches 109; Conservative 91; Mismatches 226; Indels 130; Gaps 20;

QY	34	KIVETSTSDIITGSKQPFQKLSIDELAATFSVAVETFKQFDLQKHIESIENAIDS-	92
DB	216	KIVSIEKEKID---EKSETEKLEVEIEBISCSAQVEKYKLDIAQLEENLKEKNDEISL	272
QY	93	--KLESGNVLA-----ARNNNFQPMLSPPRRNVSVETTTVSQ	130

Db 273 KOSLEDNIVLSKQVEDLVNVCQQLLEKEDKHVNRNREHNENNAEMON--LEOKFILEQ 330
Qy 131 PSOEIVPETSNGEGRMCELMCSKGLRKYIYANISDOAKLMEIIPSAKLAKPAKFAVL 190
Db 331 REHEKLOKELQIDSLLOQEKLSLHQL-----CSQCEWVKEKNLFEELKQTL 383
Qy 191 DCIGKPYL---QORRAFTKESPMSSARQVSLILLESFLLMPDRGKGVKIESWIKDEAET 247
Db 384 DELDKLQKEQEAERLVKQLEAEKSRABELKLEKL-----KGK-----EAELEK 430
Qy 248 AAVAWKRLMTEGGLAAAEKMDARGILLILVACFVPSNFRS-----TDL 291
Db 431 SSAA-----HTCATLLQEKYDS---MVSLEDVTAQFESYKALTASEIEDKLESSL 481
Qy 292 LDIRMGSN-----EIAGALKRSQFLVPMVSGIVESSIKRGHIEALEMNVYTFGMEDK 345
Db 482 QEKAAAGKAGNAEDVORQILATESNQEVYRMLDLOTKSALKETEIKEITVSFLQKITD- 540
Qy 346 FSAALVLTSLFKMSKESPERAKKQSPKAFKBAATKOLAVLSSVMQCMETHKLDPAKEL 405
Db 541 -----LQNLQKQOEEDFRKQLEDEGRKAEKENTAEI-----TEEI 577
Qy 406 PGWQ-IKEQIVSLEKD-TLQLDK-EMBEKARSLSMEBAALAKRMYN-----Q 450
Db 578 NKWRLLYEELYNKTKPFQQLDAFEVEKQA-----LLNEHGAQEQOLNKIRDSYAKLLHQ 633
Qy 451 QIKRPLSPNEMPPVTSYSSPIYDRSPPSORDDDQDEISALVSSYLQ-----PSTSEF 505
Db 634 NLKQKTKHVVLKDENSQKSEVSKURCQLANKQSETKLOBELNKVLGKIHFPDPSKAFH 693
Qy 506 HRSRRSPYVMVPLPHG 521
Db 694 HESKENFALKTPLKEG 709

RESULT 2

US-09-893-519A-73
; Sequence 73, Application US/09893519A
; Publication No. US20030027243A1
; GENERAL INFORMATION:
; APPLICANT: ANADYS PHARMACEUTICALS, INC.
; APPLICANT: THOMPSON, Craig
; APPLICANT: MOORE, Jeffrey
; APPLICANT: BUURMAN, Ed T.
; APPLICANT: BRADLEY, John
; APPLICANT: DESILVA, Tamara
; APPLICANT: HARRIS, Sandra
; APPLICANT: KOMARNITSKY, Svetlana
; APPLICANT: MENDILLO, Marc
; APPLICANT: MOORE, Daniel
; APPLICANT: MCCOY, Melissa
; APPLICANT: SANDERSON, Karen
; APPLICANT: HAQ, Tariq
; APPLICANT: ZHU, Shuhao
; APPLICANT: LONG, Fan
; APPLICANT: DAVIDOV, Eugene
; TITLE OF INVENTION: ANTIFUNGAL COMPOUNDS AND METHODS OF USE
; FILE REFERENCE: 0342/IG548-US2
; CURRENT APPLICATION NUMBER: US/09/893,519A
; CURRENT FILING DATE: 2001-06-28
; PRIOR FILING DATE: 2001-06-28
; PRIOR FILING DATE: 2000-06-29
; PRIOR FILING DATE: 2000-06-29
; PRIOR FILING DATE: 2000-08-10
; NUMBER OF SEQ ID NOS: 146
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 73
; LENGTH: 888
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Corresponds to SEQ ID NO: 146

; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: Human Genbank/BA088763
; DATABASE ENTRY DATE: 2000-01-06
; RELEVANT RESIDUES: (1)...(888)
US-09-893-519A-73

Query Match 4.3%; Score 132.5; DB 11; Length 888;
Best Local Similarity 20.0%; Pred. No. 0.014;
Matches 108; Conservative 94; Mismatches 206; Indels 131; Gaps 23;

Qy 16 ANPLQRQSQRRRELPAKIVETESTMDITIGSQOPQFLKSIDELAAFSVAVFTRQ 75
Db 410 SNGALSTEEEREEMKQM-EVYRSHSKFMKNKIGQVKQ-ELSRKDTELLALQTKLETNQ 467
Qy 76 FDLQKHIESIENAIKSDKLESNGVVLAAARNNNHQPMLSPNNVSVETTVTVSOPSOEI 135
Db 468 FSDSKQHIIEVLESUTAK-EORAAAIL-----QTEVDALRLLEKETM-UNKTKQI 517
Qy 136 VPETSNKPEGGRMCELMCSKGLRKYIYANISDOAKLMEIIPSAKLAKPAKFAVLPCIGK 195
Db 518 QDMAEEK-GTQAGEIHDLKMDLVKRVNVLQKTIENLQEQLR----- 560
Qy 196 FYLQORRAFTKESPMSSARQ-----VSLILLESFLLMPDRGKGVKIESWIK 242
Db 561 -----DREKQMSLUKERVKSLQADTTNTDTALTTLLEALAEKERTIERLK-BQDR 610
Qy 243 DEATAAVAWKRLMTEGGLAAAEKMDARGILLILVACFVPSNFRSTDLLD-----IRMS 298
Db 611 DERE-----KQBEIDN-----YKDLKOLKEVSLQGLDSEKEASLLDLKEHASSLA 658
Qy 299 GNEIAGALKRSQFLVPMVSGIVESSIKRGHIEALEMNVYTFGMEDKFSALVITS----- 354
Db 659 SDESSKQAQAEVDRLLLELKEVENEKNDKQKIAELESLSRQVKDQNKVKVANKHKEQV 718
Qy 355 FLKMSKESPERAKRK-----AQSPKAFKBAATKOL-----AVLSSVMQCM-----E 395
Db 719 EYKSAQMLEEARREDNLDSSQOLQVEELLMAWEXVKQLESMAKLSSTQOSLAEK 778
Qy 396 TH-----KLDPAKELPGWQIKQIVSLEKDTLQLDKEMEERKARSLSM-----EEA 441
Db 779 THTNLRAERRKHL-----BEVLEMKQEAALL-AAISEKDANIALLELSSSKKTKTEEV 830
Qy 442 ALAKEMYN---QOIKRPLSPNEMPPVTSYSSPIYDRSPPSORDD-----DQDE 489
Db 831 AALKREKORLVQQLAQOQTKQNRKUL-----MADNYEDDFKSHSHSNQTNHKSPOQDE 882

RESULT 3

US-10-204-887-110
; Sequence 110, Application US/10204887
; Publication No. US20030124569A1
; GENERAL INFORMATION:
; APPLICANT: INCYTE GENOMICS, INC.
; APPLICANT: PANZER, Scott R.
; APPLICANT: SPIRO, Peter A.
; APPLICANT: BANVILLE, Steven C.
; APPLICANT: SHAH, Purvi
; APPLICANT: CHALUP, Michael S.
; APPLICANT: CHEN, Simon C.
; APPLICANT: D'SA, Steven A.
; APPLICANT: AMSHEY, Stefan
; APPLICANT: DAHL, Christopher R.
; APPLICANT: DAM, Tam C.
; APPLICANT: DANIELS, Susan E.
; APPLICANT: DUFOUR, Gerard E.
; APPLICANT: FLORES, Vincent
; APPLICANT: FONG, Willy T.
; APPLICANT: GREENAWALT, Lila B.
; APPLICANT: HILLMAN, Jennifer L.
; APPLICANT: JONES, Anissa L.
; APPLICANT: LIU, Tommy F.
; APPLICANT: ROSEBERRY, Ann M.


```
; APPLICANT: ROSEN, Bruce H.
; APPLICANT: RUSSO, Frank D.
; APPLICANT: STOCKREHER, Theresa K.
; APPLICANT: DAFFO, Abel
; APPLICANT: WRIGHT, Rachel J.
; APPLICANT: YAP, Pierre E.
; APPLICANT: YU, Jimmy Y.
; APPLICANT: BRADLEY, Diana L.
; APPLICANT: BRATCHEY, Shawn R.
; APPLICANT: CHEN, Wensheng
; APPLICANT: COHEN, Howard J.
; APPLICANT: HODGSON, David M.
; APPLICANT: LINCOLN, Stephen E.
; TITLE OF INVENTION: SECRETORY MOLECULES
; CURRENT APPLICATION NUMBER: US/10/204,887
; CURRENT FILING DATE: 2002-08-21
; PRIOR APPLICATION NUMBER: 60/185,215; 60/185,216; 60/205,232; 60/205,323; 60/205,287;
; 60/205,324; 60/205,286
; PRIOR FILING DATE: 2000-02-24; 2000-02-24; 2000-05-16; 2000-05-17; 2000-05-17;
; 2000-05-17; 2000-05-17
; NUMBER OF SEQ ID NOS: 159
; SOFTWARE: PERL Program
; SEQ ID NO 110
; LENGTH: 771
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20030124569A1 LI:350243.2.orf2:2000MAY01
; US-10-204-887-110

Query Match      4.2%; Score 129; DB 15; Length 771;
Best Local Similarity 19.0%; Pred. No. 0.022;
Matches 139; Conservative 87; Mismatches 289; Indels 216; Gaps 27;

QY      2  SNYPPTAAQPTTTANPLQRHQSQRRRELPKIVET-----ESTSMDTITGCSQ 52
DB      29  SRQPPSLAASGSSGLTAIVQAAMPYSPQKIQPTAHCYGRTFHSQPCLDSSQSQEKE 88
QY      53  PQFLKSI-----DELAAFSAVETFKQFDLDQKHIESINADSKLESNGVVL 101
DB      89  RQIIEGIFKGTCKSDPDWTLVEQWSHYSPFRESSGNGMKMEGLLNGSDPHQSLASI 148
QY      102  AARNNHQPMLSPPNNVSVETTVTSQPSQETVP---ETSNKPEGGRMCELMCSKGLR 158
DB      149  KAEADKIYFTDNAPSGIGSSRLNTTPTQPTPLHVVTQNGAAS-----SVKN 201
QY      159  KYIYANISDQAK---LMSEIPS--ALKLAKEPAKFVL-----DCIGKFLYQGRRAFTKE 207
DB      202  SPAYSDISDAGEDGEGKVDYVSKSDASQLVKEGAKTLFPQPQSKDSPIYQGFESYSP 261
QY      208  S-----PMASARQVSLILLESFLMPDRGKGVKIESIKDEAETAAVAWKRLM 257
DB      262  SYAOSSPGALNPSSQAG-----VESQALKTKRDEEPESIEGKVKND----- 302
QY      258  TEGGLAAAEKMDARGLLILVACFGVPSNFRSTDLDDIIRMSGNEIAGALKRSQFLVPMV 317
DB      303  -----ICEKKPELSSSQSPSVIQORPNMYMQSLYNNQYAVVP 342
QY      318  SGIVESSIKRGHIEALEMVTTFQMEKFAALVLTSLFKMSKESFERAKR-----KAQSP 373
DB      343  YGYSDQSYH--THLLSTNTAYRQYEE-----QCKRQSLQEQQGVYDKKAENG 388
QY      374  LAFKEAATKO-----LAVLSSVMQCMET-----HKLDPAKE 404
DB      399  LKERAALKKEWKQKPSIPPTLTAKPSITDLVKGPGKAKEPAGNPAKSVIIEKLDSSK 448
QY      405  LFGMQIKQIVSLEKOTIQLDKMEEXKARSLSLMEEAALAKRMYNQOIKPRPSPVEMPP 464
DB      449  LPG-QAPEGLVKVLSADSHLSKEASEAKTGACGQAEQEMDPIILWYRQEAEPRTWTVYP- 506
QY      465  VTSSYSYSP-----YRDRSPSQEDDDQDDEI-----SALVS 495
```

```
DB      507  ---AKYSDIKSEDERWKEERDRKLEERSRSKDSVPKEDGKSTSSDCKLPTSSESRIGS 563
QY      496  SYLGPSTSP-----HRSRRSPYMWPLPHGGLGRSVYAEHLAPNSYSPGH-GHRLH 547
DB      564  KEPRPSVHVVSPLTQHQS-----YIPYMHG-----YSIS-----QSYDPNHPYSR-- 605
QY      548  ROYSPSLVHGQRHPLQYSPPIHQOQOLPYGIO-----RYVRHSPSPERYLG 593
DB      606  ---SMPAVMMQNYPGSYLPSSYSFS--PYGSKVSGEDADKARASPSVTCSSSESKALD 660
QY      594  LSNQSPRNS 604
DB      661  ILOQHASHYKS 671

RESULT 4
US-09-291-417-23
; Sequence 23, Application US/09291417A
; Publication No. US20030050230A1
; GENERAL INFORMATION:
; APPLICANT: FLOWMAN, GREGORY
; APPLICANT: MARTINEZ, RICARDO
; APPLICANT: WHYTE, DAVID
; TITLE OF INVENTION: STE20-RELATED PROTEIN KINASES
; FILE REFERENCE: 240/300
; CURRENT APPLICATION NUMBER: US/09/291,417A
; CURRENT FILING DATE: 1999-04-13
; EARLIER APPLICATION NUMBER: US 60/081,784
; EARLIER FILING DATE: 1998-04-14
; NUMBER OF SEQ ID NOS: 147
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 23
; LENGTH: 786
; TYPE: PRT
; ORGANISM: Mammalian (Human) SULU3
; US-09-291-417-23

Query Match      4.1%; Score 127.5; DB 11; Length 786;
Best Local Similarity 20.1%; Pred. No. 0.032;
Matches 142; Conservative 98; Mismatches 237; Indels 231; Gaps 33;

QY      3  NY-----PPTVAAQPTTTANPLQRHQSQRRRELPKIVETESTMDITIGOSKQPFUK 57
DB      190  NYREGDPTFRASDPQSP--PQVSRHKSHYRNREHPATIRTASL-----VTRQWQHE 240
QY      58  SIDELAAFSVAVETFKQFDLDQKHIESINADSKLESNGVVLAAARNNHQPMLSPPR 117
DB      241  QDSELREQMSGYKMRROH---QKQMTLENKLAEMDEHRLRL-----DKOLETOR 289
QY      118  NNVSVEVTVTVSQSQBIVPET---SNKPEGGRMCELMCSKGLRKYYIYANISDQAKLME 174
DB      290  NNFAAEWEKLIKXQAAMEKEAKVMSNE-----EKKFQOHI-----QAQKKE 332
QY      175  IPSAL-----KLAKEPAKFVLDCTGKFLYQGRRAFTKESPMSSARQVSLILLESFLJM 227
DB      333  LNSPFSQKREYKLRKEQLKEELN-----ENQSTPKKEKEWL----- 370
QY      228  PDRGKGVKTESWKDB-----AETAAVAWKRLM-----TEGGLAAAE----- 266
DB      371  ---SKQENICHQAEHEANLRRQYQYLECRRFKRRMLLGRHNLQODLVREELNKEQ 427
QY      267  KMDARGLLILLVACFGVPS-NFRSTDLL-----DLIRMSGNEIAGALKRSQFLVPMWSG 319
DB      428  TQKDLHAMLLRHQESMQELEFRHLNTIQMRCELIRLOHOTELTNQLEYNKR----- 481
QY      320  IVESSIKRGHIEALEMVTTFQMEKFAALVLTSLFKMSKESFERAKKAOSPLA----- 375
DB      482  --ERELRR-KHV-----NEVROQPKSLAKSKELQIKKQFQDTCKLTQRYKALRNH 528
QY      376  -----FKEAATKQLAVLS-----SVMQCMETH--KLDPAKELPGHQIKE 412
DB      529  LLETPPKSEHKAVLKRKEEQTKALILAEQYDHSINEMLSQALRLDEAQAECQVLKY 588
```

QY 413 QI-----VSLEKDTLQDKEMEKEKARSLSL-----MEEAALAKRMVNOQ 451
Db 589 QI-----VSLEKDTLQDKEMEKEKARSLSL-----MEEAALAKRMVNOQ 451
QY 452 IKRPLSPMEMPVTSSSYPIYDRSPSPQDDDDQDEISALVSSYLGSTSPFHSRRS 511
Db 646 TERIR-----SLLERQAREIEAFDSBSMRGLGFSNMVLSNLS 896
QY 512 PEYVPLPHGLGRSVYAYEHLAPN-----SYSPGHG---HRLHRQYSPSLVHGQHPLOY 564
Db 682 PE-----AFSHSYPGASGWSHNPTGPGPHGHPMGPPQAWG--HPMQG 724
QY 565 SP-----PIHQQQQLPYGIQVYRHSPEERYL-----GLSNQSPRNS 604
Db 725 GPQPMGHPSPGMPGVPRGSSMGVNSPQALRRTASGGRTEQGMRSRST 772

RESULT 5
US-09-291-417-31
; Sequence 31, Application US/09291417A
; Publication No. US20030050230A1
; GENERAL INFORMATION:
; APPLICANT: PLOWMAN, GREGORY
; APPLICANT: MARTINEZ, RICARDO
; APPLICANT: WHYTE, DAVID
; TITLE OF INVENTION: STE20-RELATED PROTEIN KINASES
; FILE REFERENCE: 240/300
; CURRENT APPLICATION NUMBER: US/09/291,417A
; CURRENT FILING DATE: 1999-04-13
; EARLIER APPLICATION NUMBER: US 60/081,784
; EARLIER FILING DATE: 1998-04-14
; NUMBER OF SEQ ID NOS: 147
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 31
; LENGTH: 1001
; TYPE: PRT
; ORGANISM: Mammalian (murine/human) SULU3
US-09-291-417-31

Query Match 4.1%; Score 127.5; DB 11; Length 1001;
Best Local Similarity 20.1%; Pred. No. 0.047;
Matches 142; Conservative 98; Mismatches 237; Indels 231; Gaps 33;

QY 3 NY-----PPTVAAQPTTANPLLRHQSRRELPKIVETESTSMDITIGSQKQPFLLK 57
Db 405 NYREGDPRTASDPQSP--PQVSRHKSHYRNREHFATIRTA--VTRQMOEHE 455
QY 58 SIDELAASFVAVETFKROFDDLKQKHIESTENALDSKLSNGVVLAAARNNPHQPMSPR 117
Db 456 QDSELREQSGYKRMRRQH--QKQMTLLENKKAEMDEHRLRL-----DKDLETQR 504
QY 118 NVSVETTTVTSQSOEIVPET---SNKPEGGRMCMLCCKGLRKYIYANISDQAKLME 174
Db 505 NNFAAEMEKLIKQHAAMEKEAKWMSNE-----EKKFQOHI-----QAQKKE 547
QY 175 IPSAL-----KLAKPAKFLVDCIGKPYLQGRAPFTKESPMSSARQVSLILLESFLLM 227
Db 548 LNSFLESQREYKLRKEQLKBEIN-----ENQSTPKKEQEWL----- 585
QY 228 PDRGKGKVKIESWIKDE-----AETAAWVRKFLM-----TEGGLAAAB----- 266
Db 586 ---SKQENIQHFQABEENALLRRQYLECECRFKRMMLGRNLEQDLVRELNKRQ 642
QY 267 -KYDAGLLLVACFGVPS-NFRSTDLL-----DLIRMSGNEIAGALKRSQFLVPMVSG 319
Db 643 TOKDLEHALLRQHESMQELFRHLNANTIOKRCELIRLOHQTFLTNQLEYNKR----- 696
QY 320 IVESSTIKRGMHIBALEMVYTFMGEDFSAALVLTSLFKMSKESFERAKKQASPLA----- 375
Db 697 --FRELR-KHV-----NEVQQPKSLKSELQIKKQFQDTCKIQTOYKALENH 743
QY 376 -----FKEAATKQLAVLS-----SVNQCWNETH--KLDPAKELPGWQIKE 412

Db 744 LLETTPKSEKAVLKUKESQTRKLAIALAQYDHSINEMLSLQALRLDEQAQAECCVLKM 803
QY 413 QI-----VSLEKDTLQDKEMEKEKARSLSL-----MEEAALAKRMVNOQ 451
Db 804 QIQOELELLNAYQSKIKMQAEA-QHDERELRELEQVSLRRALLQEKIEEMLA--LQNER 860
QY 452 IKRPLSPMEMPVTSSSYPIYDRSPSPQDDDDQDEISALVSSYLGSTSPFHSRRS 511
Db 861 TERIR-----SLLERQAREIEAFDSBSMRGLGFSNMVLSNLS 896
QY 512 PEYVPLPHGLGRSVYAYEHLAPN-----SYSPGHG---HRLHRQYSPSLVHGQHPLOY 564
Db 897 PE-----AFSHSYPGASGWSHNPTGPGPHGHPMGPPQAWG--HPMQG 939
QY 565 SP-----PIHQQQQLPYGIQVYRHSPEERYL-----GLSNQSPRNS 604
Db 940 GPQPMGHPSPGMPGVPRGSSMGVNSPQALRRTASGGRTEQGMRSRST 987

RESULT 6
US-10-177-293-254
; Sequence 254, Application US/10177293
; Publication No. US20030124129A1
; GENERAL INFORMATION:
; APPLICANT: Glatt, James
; APPLICANT: Lillie, James
; APPLICANT: Zhao, Xumei
; APPLICANT: Gannavarpu, Manjula
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Mertens, Maureen
; APPLICANT: Myer, Vic
; APPLICANT: Wang, Youzhen
; APPLICANT: Xu, Yongyao
; APPLICANT: Hoersch, Sebastian
; APPLICANT: Monahan, John
; APPLICANT: Meyers, Rachel E.
; APPLICANT: Bast Jr., Robert C.
; APPLICANT: Hortobagyi, Gabriel N.
; APPLICANT: Pusztai, Lajos
; APPLICANT: Meric, Funda
; APPLICANT: Sahin, Aysegul
; APPLICANT: Mills, Gordon B.
; TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR IDENTIFICATION, ASSESSMENT,
; FILE REFERENCE: MRI-038
; TITLE OF INVENTION: PREVENTION, AND THERAPY OF BREAST CANCER
; CURRENT APPLICATION NUMBER: US/10/177,293
; CURRENT FILING DATE: 2002-06-21
; PRIOR APPLICATION NUMBER: US 60/299,887
; PRIOR FILING DATE: 2001-06-21
; PRIOR APPLICATION NUMBER: US 60/301,572
; PRIOR FILING DATE: 2001-06-27
; PRIOR APPLICATION NUMBER: US 60/306,501
; PRIOR FILING DATE: 2001-07-18
; PRIOR APPLICATION NUMBER: US 60/325,002
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US 60/362,585
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: US 60/xxx,xxx
; PRIOR FILING DATE: 2002-05-14
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 254
; LENGTH: 1005
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-177-293-254

Query Match 4.1%; Score 127.5; DB 15; Length 1005;
Best Local Similarity 20.1%; Pred. No. 0.047;
Matches 142; Conservative 98; Mismatches 237; Indels 231; Gaps 33;
QY 3 NY-----PPTVAAQPTTANPLLRHQSRRELPKIVETESTSMDITIGSQKQPFLLK 57

Db 409 NYREGDPRTRAGDPQSP--PQVSRHSHYRNREHFATIRTASL-----VTRQMOEHE 459
QY 58 SIDELAAPFVAVETPKRQDDLOKHTESTENAIKSLNSGVVLAARNNNHQPMLSPPR 117
Db 460 QDELEQWGYKRMRRQH--OKQMTLENKKAEMDEHRL-----DKOLETOR 508
QY 118 NNYSVETTTVSQFQIEIVPET---SNKPEGGRMCELMCSKGLRKYIYANISDOAKLME 174
Db 509 NNPAEAMEKIKKHOAMEKEAKVMSNE-----EKKFQCHI-----QAQOKKE 551
QY 175 IPSAL-----KLAPEKAFVLDICGKYFLOGRAFTKESPMSSARQSVLLILLESFLM 227
Db 552 LNSFLESQREYKLRKEQLKELN-----ENQSTPKKEKEWL-----589
QY 228 PDGKGVKTESWIKDE-----AETAAVAWRKELM-----TEGGAAAE-----266
Db 590 ---SKQENIQHQAEEANLLRRQRYLELECRFKRMLLGRHLEODLAVRELNKRQ 646
QY 267 -KMDARGLLLLVACFGVPS-NFRSTDLL-----DLIRMSGSEIAGALKRQSVLPVMSG 319
Db 647 TQXDLPHAMLLRQHESMQBLEPHNLTIQKRCLEJIRLOHOTELTNQLEYNKR-----700
QY 320 IVBSS:KRGHMBIALEWVTFGMEDKFSAAVLVTSFLKMSKESFERAKRKAQSPILA-----375
Db 701 --BRELR-KHV-----NEVROQPKSLSKELQIKQFQDTCKIOTROYKALRNH 747
QY 376 -----FKBAATQKLAVLG-----SYMOCMETH--KLDPKELFGWQIKE 412
Db 748 LLETTPKSEHKAVALKRLKEEOTRKLAILAEQDHSINEMLSLQALRLDEAQEAECQVLKM 807
QY 413 QI-----VSLKDTLOLDKEMEERKARSLSL-----MEEAALAKEMYNQO 451
Db 808 QLOQELELLNAYOSKIKWQAEA--QHDRELELEORVSLRALLLEOKIEBMLA--LQNER 864
QY 452 IKRPR:SPMEMPPVTSSSYPIYDRSPSQDDQDEISALVSSYLGPFSTSPFHRSSRS 511
Db 865 TERIR-----SLLERQAREIEAFDSSEMLGFSNNVLSNLS 900
QY 512 PEYMWVLPHGGLGRSVYAYEHLAPN---SYSPGHG---HRLHROYSPSLVHGQHPLOY 564
Db 901 PE-----AFSHSYFGASGWSHNPATGGPGHWHGPPQAWG--HPMQG 943
QY 565 SP-----PIHQOOLPYGIQVRYRHSRERYL---GLSNQSRPSNS 604
Db 944 GPQPGWGHSPGMGVPRGSSMGVNSPQALRRTASGRTEQGMRSST 991

RESULT 7

US-09-919-172-98
; Sequence 98, Application US/09919172
; Patent No. US20020119463A1
; GENERAL INFORMATION:
; APPLICANT: Paris, Mary
; APPLICANT: Turner, Christopher M.
; TITLE OF INVENTION: PROSTATE CANCER MARKERS
; FILE REFERENCE: PA-0036 US
; CURRENT APPLICATION NUMBER: US/09/919,172
; PRIOR FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: 60/222,469
; PRIOR FILING DATE: 2000-07-28
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: PERL Program
; SEQ ID NO 98
; LENGTH: 3256
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20020119463A1 2700132CD1
US-09-919-172-98

Query Match

4.1%; Score 127.5; DB 10; Length 3256;

Best Local Similarity 20.0%; Pred. No. 0.31;
Matches 137; Conservative 92; Mismatches 228; Indels 229; Gaps 32;
QY 11 QTTTANPLLQHQHQEQRR--RELPKIVETESTMDI-----TIGSQKQPOFLK 57
Db 864 ESKTVSTVNRGREGTEPRNIQKUPVESKSEETTEIVECILKRGQKATLLQQRSEGMK 923
QY 58 SIDELAAPFVAVETPKRQDDLOKHIESIENAIKSLNSGVVLAARNNNFHQ---PMLS 114
Db 924 EIE-----RPFETVKENIELKEN--DEKMK-----MKSRWTGOKCAPM-- 961
QY 115 PRNNVSVEITTVSQFQIEIVPETS-----KPEGGRMCELMCSKGLRKY 160
Db 962 -----SDLTDLKSLPDTLMKDTARQONLLOTHAKAPKSEKGIKTMPCOSLOPEP 1014
QY 161 IVANISDOAKLMEETPSALKLAKPAKVLDCIGKF-YLOGRAFTKESPMSSARQSVLL 219
Db 1015 INTPTHTKQOL-----KASLGKGVKEELLAVGKFTRTSGTTHTHREAPAGDKSI--- 1065
QY 220 ILESFLLMPD-----RGKGVKIESWIKDEATAA--VAVRKELMTEGGIAAEKMDAR 271
Db 1066 --RTFKESFKQILDPAARVTGMKKWPTPKKEAQSLDLGAFKELFQTPGPSEESMTDEK 1123
QY 272 GLLLLVACFGVP-----SNFRSTDL-----LDLIRMSGSEIAGALKRSQF 312
Db 1124 --TTKIACKSPPEVSDTPTSTKQWPKESLRKADVEBEFLALRLTFS---AG---KAM 1174
QY 313 LPMVSGIVSESIK--RGMHIEALEMVTTFQMEDKFSAAVLVTSFLKMSKESFERAKRKA 370
Db 1175 LTPKAGGDEKDIKAFMGTPVQKLDLAGT-----LPGSKRQLQTPKEKA 1218
QY 371 QS--PLA-FKE-----AATKO-----383
Db 1219 QALEDLGAFKELFQTPGHTTELVAAAGTKTKIPCDSQSDPDVDTSTKQPKRSIRKADV 1278
QY 384 ----LA-----VLSSVMQCMET-----HKLDPAKELPGWQIKEIVSL 417
Db 1279 EGBELLACRLMPSAGKAMHTPKPSVGEBKDIIFVGTVPVQKLDLTENLTGSKRRPQTPKE 1338
QY 418 EXDTLOLDKEMEERKARSLSLMEEAALAKEMYNQIKRPLSPMEMPPVTSSSYPIYDR 477
Db 1339 EQALEDUTGFKELFQTPGHTTEEAAG-----KTKMPCESPPESADPTSTTRQOP 1391
QY 478 SPFSQRDDQDEISAL--VSSYLGPFSTSPFHRSSRSPYVMVPLPHGGLGRSVYAYEHLAP 535
Db 1392 KTFLEKRDVQKELSALKLTQTSGETT--HTDK-----VP---GGEDKSINAFRETAK 1439
QY 536 NSYSPGHGHLHRLHROYSPSLVHGQHP 561
Db 1440 QKLDP-----AASVTGSKRHP 1455

RESULT 8

US-09-919-039-21
; Sequence 21, Application US/09919039
; Publication No. US20030108871A1
; GENERAL INFORMATION:
; APPLICANT: Kaser, Matthew R.
; TITLE OF INVENTION: GENES EXPRESSED IN TREATED HUMAN C3A LIVER CELL CULTURES
; FILE REFERENCE: PA-0035 US
; CURRENT APPLICATION NUMBER: US/09/919,039
; CURRENT FILING DATE: 2002-09-09
; PRIOR APPLICATION NUMBER: 60/222,113
; PRIOR FILING DATE: 2000-07-28
; NUMBER OF SEQ ID NOS: 401
; SOFTWARE: PERL Program
; SEQ ID NO 21
; LENGTH: 3256
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20030108871A1 2700132CD1

US-09-919-039-21

```
Query Match      4.1%; Score 127.5; DB 11; Length 3256;
Best Local Similarity 20.0%; Pred. No. 0.31;
Matches 137; Conservative 92; Mismatches 228; Indels 229; Gaps 32;

QY 11 QPTTANPLLRHQGEQR--RELPKIVETSTSDI-----TIQSQKQPOFLK 57
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 864 EPSKTVSTVNRSGRSTEFENIKLQPVESKEETNTEIVCEILKRGOKATLLQORREGMK 923
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 58 SIDELAARFSAVETFRQPDLDQKIESIENAIKSLKESNGVLAARNNFHQ---PWLS 114
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 924 EIZ-----RPETTKENIELKEN--DEKMA-----MKSRWTGQKCAPM-- 961
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 115 PRNNVSVETTVTSOPSOEIVPETS-----KPEGRCMELCMCKGLRKY 160
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 962 -----SDLTDLKSLPDTLTKMKTARGONLLQTOCHAKAPKSKGKITKMFQCSLOPEP 1014
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 161 IYANISDQAKMEEIPSAKLAKEPAKFLVDCIGKF--VLQGRRAFTKESPMSSARQVSL 219
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1015 INPPTHKQOL-----KASLGKVGKVELLAVGFTTSGETTHREPAAGDKSI--- 1065
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 220 ILSEFLMPD-----RGKGKVKIESWIKDAETAA--VAMKRLMTBGGAAAEKVDAR 271
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1066 --RTFKESPQILDPAARVTGMKWPRTPKERAQSLDLGAFKELFQTPGPSSESMTDEK 1123
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 272 GULLLVACFVP-----SNFRSTDL-----LDLRMSGNEIAGALKRSQF 312
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1124 --TTKIACKSPPPESVDTPTSTKQMPKRSKADVEEBFLALRLKLTPS---AG----KAM 1174
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 313 LVPWMSGIVESSIK--RGNHIBALEMVTYTFGMEDKFSALVLTSLFKMKSSEFRACKKA 370
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1175 LTPKAGGDEKOIKAFMGTFVQKLDLAGT-----LPGSKQLQTPKEKA 1218
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 371 QS--PLA-FKE-----AATKO----- 383
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1219 QALDLGAFKELFQTPGHTTELVAAAGTKTKIPCDSPQSDPVDPTPTSTKPKRSIRKADV 1278
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 384 ----LA-----VLSSVMOCMET-----HKLDPAKELFGWQIKQIVSL 417
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1279 EGELLACRLNMPGAKAMHTPKPSVGEKDIIFVGTPTVQKLDLTENLTGSKRRPQTPKE 1338
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 418 EKDTLQDKMEMEKARSLIMEEALAKMWNQOIKRPLSPMEMPPVTSSSYPIYDR 477
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1339 EAQALDLGTFKELFQTPGHTTEAVAAG-----KTKMPCSESSPDSADPTPTSTRQP 1391
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 478 SPFSQDDQDDETSAL--VSSVLGPSTSPHRSRSPENWVPLPHGLGRSVYAVEHLAP 535
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1392 KTFLEKRDVQKELSALKLTQTSGETT---HTDK-----VP---GGEDKSIINAFRETAK 1439
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 536 NSYSPGHGHLRHQVSPSLVHGQRHP 561
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1440 QKLDP-----AASVTGSKRHP 1455
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

RESULT 9

```
US-10-097-340-43
; Sequence 43, Application US/10097340
; Publication No. US20030087250A1
; GENERAL INFORMATION:
; APPLICANT: John MONAHAN
; APPLICANT: Manjula GANNAVARAPU
; APPLICANT: Sebastian HOERSCH
; APPLICANT: Shubhangi KAMATKAR
; APPLICANT: Steve G. KOVATS
; APPLICANT: Rachel E. MEYERS
; APPLICANT: Michael MORRISSEY
; APPLICANT: Peter OLANDT
; APPLICANT: Ami SEN
; APPLICANT: Peter VEIBY
; APPLICANT: Gordon B. MILLS
; APPLICANT: Robert C. BAST, Jr.
; APPLICANT: Karen LU
```

```
; APPLICANT: Rosemarie SCHMANDT
; APPLICANT: Xumei ZHAO
; APPLICANT: Karen GLATT
; TITLE OF INVENTION: Nucleic Acid Molecules and Proteins For The Identification,
; TITLE OF INVENTION: Assessment, Prevention, and Therapy of Ovarian Cancer
; FILE REFERENCE: MRI-030
; CURRENT APPLICATION NUMBER: US/10/097,340
; CURRENT FILING DATE: 2002-03-14
; PRIOR APPLICATION NUMBER: 60/276,025
; PRIOR FILING DATE: 2001-03-14
; PRIOR APPLICATION NUMBER: 60/325,149
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: 60/276,026
; PRIOR FILING DATE: 2001-03-14
; PRIOR APPLICATION NUMBER: 60/324,967
; PRIOR FILING DATE: 2001/09/26
; PRIOR APPLICATION NUMBER: 60/311,732
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: 60/325,102
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: 60/323,580
; PRIOR FILING DATE: 2001-09-19
; NUMBER OF SEQ ID NOS: 363
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 43
; LENGTH: 1203
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-097-340-43
```

```
Query Match      4.0%; Score 124.5; DB 15; Length 1203;
Best Local Similarity 19.9%; Pred. No. 0.12;
Matches 111; Conservative 91; Mismatches 213; Indels 143; Gaps 25;
```

```
QY 7 TVAAQPTTANPLLRHQHS-----EORRR-----ELPKIV----- 36
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 169 TIDTAPLSVSLINKPDSQGGQARGTRGRTTMLPPEQRKSKSLDSLPRDTEERE 228
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 37 -----ETESTSMITIGQSKQPOFLKSIDELAAPFSAVAVETFRQPD--LQKHIESIEN 88
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 229 RSTNHTSTTKYDNHVGTSKQPAQSNLSPLSGFSRS-----RQTDWVLQSFEPPRS 283
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 89 AID-SKLSNGVLAARNNNHOPMLSPRNNVSVETTV--TVSQSQEIVPETSNNKPEG 145
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 284 ADQPTMLQFKSTPDLRLD-----QGEAAPGSDVDMKATIIYIGILREGSSE--SITSVRKV 337
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 146 GRMCE-----LMCKSLKRYIYA-----NISD-QAKLMEEI-----PSALKLAKEPA 186
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 338 SIIVLEKMPVLWVSSGSTKAVAGQELTRKVEELQRLDDEVKROKLEPSQVGLERQLE 397
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 187 KPVLDICIKFYIQGRRAFTKESPMSSARQVSLILLESFLMPDEGKGVKIESWIKDEAE 246
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 398 EXTEBCSLQLELLERR---KGEAQSNKE-----LQNNKRLDQGED-----LRHGLE 442
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 247 TAAVAVRKRRLMTEGGGLAAAEKMDARGLLLVACFGVPSNFRSTDLDLIRMSGNSNIAGA 306
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 443 TQVMELQNKLKHVQGPPEPAKEVLLKDLL-----ETRELLLEV-LECKQVVEEQ 489
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 307 LK-RSQFLVPMVWSGIVESSIKRGHMEIEALEMVTYTFGMEDKFSALVLTSLFKMKSSEFR 365
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 490 LRLRERELTALKGALKEEVAQRDQVEHVROQY-----QRDTQQLRR 531
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 366 AKRKAQSPFAFKEATKOLAVLSSVMOCMETHKLDPAKELPG-WQ-----IKE 412
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 532 SNQDATQDHAVLEAERQKMSALVGLQ-----RELEETSEETGHQSMFQKNKEDLRATQ 587
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 413 QIVSLKDTLQDKMEMEKARSLIMEEALAKMWNQOIKRPLSPME-----MPP 464
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 588 ELLQLRMEKEEMEELGKIEVLQRELEQARASAGDTRQVEVLKELLRTQEBELQAE 647
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 465 VTSSSYSPYIYDRSPSPSQ 482
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 648 ROSQEVAGRHRRDRELEKQ 665
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

```

; FILE REFERENCE: GC394C1-US
; CURRENT APPLICATION NUMBER: US/09/932,183A
; CURRENT FILING DATE: 2001-08-17
; PRIOR APPLICATION NUMBER: US 09/308,375
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: PCT/US98/18928
; PRIOR FILING DATE: 1998-09-08
; PRIOR APPLICATION NUMBER: EP9719636.4
; PRIOR FILING DATE: 1997-09-15
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 2285
; TYPE: PRT
; ORGANISM: Bacillus subtilis
US-09-932-183A-2

```

```

Query Match      3.9%; Score 122.5; DB 10; Length 2285;
Best Local Similarity 19.3%; Pred. No. 0.49;
Matches 83; Conservative 70; Mismatches 15; Indels 125; Gaps 15;

```

```

QY 159 KTYANISD---QAKMBEIPSAKLAKPAKFLVDCIGKFKYLOQRRAFTKESPMSSARQ 215
Db 526 KTFARIGNNQSISIKALEQIGISVKTAGEAKSADLISE-----VAGKWDTLSDAQK 578
QY 216 VSLILLESFLMPDRGKGKVKIESWIKDEAETAA-----VAVRKRL----- 256
Db 579 QNTSIGVAGIYQLSRFNAMNFIQNAAKTAANSTGSAWSEQQKYADSLQARVKNLQN 638
QY 257 -MTEGGLAAEKMDARGLLLVACFGVPSNFRSTDLDLIRMSG--NEIAGALKRSOFL 313
Db 639 NTFEFAIADSA-----FISDGLIEFTQAAGSLLNASTGVKISGVGL 680
QY 314 VPMVSGI-----VSSIKRGHIEALEMVYTFGMEKFSALV----- 351
Db 681 PPLAAVSTATLLSKNTRTLASSLILGTFRAMQETLATAGLEAGMTAAVASRVLKTAL 740
QY 352 -----LTSP--LYMSKESFEERAKR-KAOSPLAFKEATKOLA 385
Db 741 RGLVSTLVGGAPALGWALESLSISFAEAKKAKODFQSQQTVEAITTKDSTDKLIQ 800
QY 386 VLSSVMQCMETHKLDPAKELPGWIKQIV-----SLKDTLQDKEMEE 430
Db 801 QYKELQVKRESRLTSDSEQEVYLTQQLAQTFPALVKYDQSGNAILTKNKELEKAIEN 860
QY 431 KARSLSI--MEEAALAKRMVNOQIKRPLSPNEMPPVTSYSSYSP--YRDRSFPQS---R 483
Db 861 TKYALAKKQETRDSAKTTFEDASKEIKSKDEL-----KOYKQIADYNDKGRPKWDLIA 915
QY 484 DDDODEISA 492
Db 916 DDDDYKVA 924

```

```

RESULT 12
US-10-128-714-3240
; Sequence 3240, Application US/10128714
; Publication No. US20030119013A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Bo
; APPLICANT: Tishkoff, Daniel
; APPLICANT: Hu, Wenqi
; APPLICANT: Zamudio, Carlos
; APPLICANT: Eroszhkin, Alexey M
; APPLICANT: Lemieux, Sebastien M
; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
; TITLE OF INVENTION: Methods of Use
; FILE REFERENCE: 10182-018-999
; CURRENT APPLICATION NUMBER: US/10/128,714
; CURRENT FILING DATE: 2002-04-23
; PRIOR APPLICATION NUMBER: US 60/285,697
; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: US 60/287,066

```

```

RESULT 10
US-10-156-761-12795
; Sequence 12795, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 12795
; LENGTH: 426
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-12795

```

```

Query Match      4.0%; Score 124; DB 15; Length 426;
Best Local Similarity 22.1%; Pred. No. 0.024;
Matches 95; Conservative 44; Mismatches 155; Indels 136; Gaps 20;

```

```

QY 173 BEIPSALKLA-KEPAKFLVDCIGK-----YLQRRAT--KESPMSSARQVSLILLES 223
Db 106 EEALDALLRAHEGARIVGLCTGAFVLAAGLLDGRPATTHMYATLAKRPSVHVDP 165
QY 224 FLMPDRGKGKVKIESWIKDEAETAAVAVRKRLMTEGGLAAAEKMDARGLLLVACFGVP 283
Db 166 ELFVDD---GDV-----LTSAGTAA----- 182
QY 284 SNFRSTDL-LDLIRMSGNIEAGALKRSOFLVPMVSG-----IVESSIKRGHIEALEMVY 338
Db 183 ----GIDCLHLIVTRDGNFAAGALARLVVPRRGGQERYLDRLSPBEIGADPLAEVV 238
QY 339 TFGME---DKFSAALVTSFLKMSKESFERAKR--AQSPPLAFKEATKQLAVLSSVMQC 393
Db 239 AWALEHLHEQPDVE-TLAAAYMSRRTFDRFRSLTGSAPLQ----- 279
QY 394 METHKLDPAKELPGWIKQIVSLEKDTLQDKEMEKARSLSLMBEALAKRMVNOQIK 453
Db 280 -----WLITQVLOAQRLLETSDYSVDEVAGRCGRSPVALRGHFRRLGS 325
QY 454 PRLSPNEMPPVTSYSSYSPYVDRSFPQSDDDODEISALVSLVGLPSTSFPHRSRRSPE 513
Db 326 SP-----AAVTAAYARRPOGERPVDP-----SVVGPPPALPQ-----E 360
QY 514 YMVPLPHGGLGRSVYAYEHLAPN---SYSPG-HG-HRLHROYSPSLVHGQRHPLQYS--- 565
Db 361 APVPMQ---TERTAAASGLPAASLSTPEPKGKGSHGSHGSHGSHGSHGSELYAAGR 416
QY 566 PPIHQQQQLP 575
Db 417 FSLPGQRSAP 426

```

```

RESULT 11
US-09-932-183A-2
; Sequence 2, Application US/09932183A
; Patent No. US20020127641A1
; GENERAL INFORMATION:
; APPLICANT: Estell, David A.
; TITLE OF INVENTION: Proteases From Gram-Positive Organisms

```

```

; PRIOR FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: US 60/295,890
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/303,899
; PRIOR FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: US 60/316,362
; PRIOR FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 8603
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 3240
; LENGTH: 1001
; TYPE: PRT
; ORGANISM: Aspergillus fumigatus
; US-10-128-714-3240

Query Match 3.7%; Score 115.5; DB 15; Length 1001;
Best Local Similarity 19.1%; Pred. No. 0.54;
Matches 113; Conservative 103; Mismatches 254; Indels 123; Gaps 24;

QY 7 TVAAOPTTANPLQHQSEQ--RRRLPKI-----VET-----ESTSMITIG 48
Db 93 TASTETASKKELPAKESEEFSDNELPRLSELKKEQBEVETLKSQVESLKDRLAVA 152
QY 49 QSKQPQFLKSIDELAAFSVAV-----ETFKQFDDLOKHIESINADISK 93
Db 153 RESTEGWQNLSEATRELVELDIRDKQSEIKKLDTQAEVDDVRAKLAESNTIAXA 212
QY 94 LESGVVLAARNNNHQPMLSPRNNVSEITTVVSQPSQEIVPETSNKPPGGRMCLMC 153
Db 213 GEEAEKLAELKQKTEE--IEKLQDVQVQCKNDQDKELQKLEEVAKEDASEK----- 265
QY 154 SKGLRYIYANISDQAKLMEEIPSALKAKEPAKPVLDICIKF-----YLOGR----- 201
Db 266 KLGVLQGLVDNLRSLQKDTVEVVSDI---KADMKGVBELGKLQNVVDYLNNLKDNAEV 322
QY 202 -----PAFTKESPMSSARQVSLILLESFLMMDRCKGVKIESWIKDEAEV---AAVAMR 253
Db 323 QOTRDOAAGSQSPDFSLRKSAPLQQ-----EPGGGNLATEPEPPANGATGGAGAKKK 377
QY 254 KRLMTGGGLAAAFKMDARGLLIL--VACFGVPSNFRSTDLID-----LIRMGSGNE-- 302
Db 378 KNKKKGGKGGDGTAKATGAADVQDQKSHAESDQAASDLADLEQIKIQLTKQLGDKAEA 437
QY 303 ---IAGALKRSQFLVPMVSGIVESSIKRGM-HIEALEMVIYFGMEDKF-----SA 348
Db 438 IDRLSKLKGEEGLKEBIEISLDDLLHGLQDHYEAKOKIKELNVKKALETVSKLEKEL 497
QY 349 ALVLTFLKMS-----KESFERAKRKAQSPLAFKAEATKQAV--LSSVMQCMET 396
Db 498 ADIRTSKASADSCKMHSDLKEDYENLVKLTN-LETELSAAQQLAATRFKDLTELRET 556
QY 397 -HKLDPAKELPGHQI-----KEQIVSLEKDTQLDKEMEE-----KARSLSLMEEAA 442
Db 557 LQKLQV--ELKSURVSSSELKSTKEALASKESBELRTLEGKHEELAEVTKLTISTISERDA 614
QY 443 LAKRWYNOQIKRPLSPMEMPPTSSSYPIYDRSFPQSRDDQDEISALVS 495
Db 615 EVKTL-NQKIRQETDRLKABENLTVAQSDLRVSESCKQEAETVEKIAADLS 666

RESULT 13
US-10-177-293-391
; Sequence 391, Application US/10177293
; Publication No. US20030124128A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Glatt, Karen
; APPLICANT: Zhao, Xumei
; APPLICANT: Gannavarpu, Manjula
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Mertens, Maureen
; APPLICANT: Myer, Vic
; APPLICANT: Wang, Youzhen
; APPLICANT: Xu, Yongyao
; APPLICANT: Hoersch, Sebastian
; APPLICANT: Monahan, John
; APPLICANT: Meyers, Rachel B.
; APPLICANT: Bast Jr., Robert C.
; APPLICANT: Hortobagyi, Gabriel N.
; APPLICANT: Pusztai, Lajos
; APPLICANT: Meric, Funda
; APPLICANT: Sahin, Aysegul
; APPLICANT: Mills, Gordon B.
; TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR IDENTIFICATION, ASSESSMENT,
; TITLE OF INVENTION: PREVENTION, AND THERAPY OF BREAST CANCER
; FILE REFERENCE: MRI-038
; CURRENT APPLICATION NUMBER: US/10/177,293
; PRIOR FILING DATE: 2002-06-21
; PRIOR APPLICATION NUMBER: US 60/299,887
; PRIOR FILING DATE: 2001-06-21
; PRIOR APPLICATION NUMBER: US 60/301,572
; PRIOR FILING DATE: 2001-06-27
; PRIOR APPLICATION NUMBER: US 60/306,501
; PRIOR FILING DATE: 2001-07-18
; PRIOR APPLICATION NUMBER: US 60/325,002
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US 60/362,585
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: US 60/xxx,xxx
; PRIOR FILING DATE: 2002-05-14
; NUMBER OF SEQ ID NOS: 506
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 391
; LENGTH: 862
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-177-293-391

Query Match 3.7%; Score 114.5; DB 15; Length 862;
Best Local Similarity 18.9%; Pred. No. 0.53;
Matches 88; Conservative 97; Mismatches 184; Indels 97; Gaps 20;

QY 28 RRRLPKIVETESMTDITIGSKQPQFLKSIDELAAFS--VAVETFKQFDD----- 78
Db 403 RRAQSTDSLGTSGLSQSKALGYNKAKSAGNLDE--SDFGPLVGADSVSENFDTASLSLQ 461
QY 79 -----LOKHIESINADISKLESNGVVLAARNNNHQPMLSPRNNVSEITTVVSQPS 132
Db 462 MPSGFWLTKQERAIKANTPQEBETASLLSVTQGMESAYVPSGYELVSETENWLLQKE 521
QY 133 QSIVPETSNKPPGGRMCLMCS---KGLR--KIYANISDQAKLM-----EEIPSAL 179
Db 522 ---VHNAGNKL--GERCD-MCSNYEKQLQGIQIEAETRDQVKKLQMLRQANDQLEKTM 575
QY 180 KLAKPAKPVLDICIKFYLQGRRAFTKESPMSSARQVSLILIE---SFLMPDRGKGVK 236
Db 576 KQKLELEDFI-----KQSEDSHQISALVLRQAQASEILLEELQOGLSQ 619
QY 237 IPSWIKDEAETAAVNRKRLMTEGGAAAEKNDARGLLLVACFGVPSNFRSTDLILR 296
Db 620 AK---RDVQEQMAVLMQSREQVSEELVRLQK-----DNDS 651
QY 297 MGSNEIAGALKESQ-FLVPMVSGIVESSIKRGMHIEALEMVIYTFG--MEDKPSAALVLT 353
Db 652 LQKSHLVSLQQAEDFILDPTTEALRELVLK--YREDINVRTAADHVEKLAEL-- 707
QY 354 SFLKMSKESFERAKRKAQSPLAFK-EAATKQAVLSSVMQCMETHKLDPAK-ELPGWQIK 411
Db 708 -FLKEQIQAEQCLKENLEETLQLEIENCKEBIASISLKABLERIKVEKQGLESTLREKS 766
QY 412 EQIVSLEKDTQLDKEMEEKARSLMEEAALAKRWYNOQIKRPL 457
Db 767 QQLESQLEIKISLEBQLKKEATAKAIVFQI-----MPEEKNAQRL 807

RESULT 14
```

```
US-10-171-311-164
; Sequence 164, Application US/10171311
; Publication No. US20030087270A1
; GENERAL INFORMATION:
; APPLICANT: Schlegel, Robert
; APPLICANT: Chen, Yan
; APPLICANT: Zhao, Xumei
; APPLICANT: Monahan, John
; APPLICANT: Kamathkar, Shubhang
; APPLICANT: Glatt, Karen
; APPLICANT: Ganavarsapu, Manjula
; APPLICANT: Hoersb, Sebastian
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY
; TITLE OF INVENTION: OF CERVICAL CANCER
; FILE REFERENCE: MRI-035
; CURRENT APPLICATION NUMBER: US/10/171,311
; CURRENT FILING DATE: 2002-06-12
; PRIOR APPLICATION NUMBER: US 60/298,159
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,155
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/335,936
; PRIOR FILING DATE: 2001-11-14
; NUMBER OF SEQ ID NOS: 238
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 164
; LENGTH: 1938
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-171-311-164

Query Match 3.7%; Score 114.5; DB 15; Length 1938;
Best Local Similarity 19.7%; Pred. No. 1.9;
Matches 124; Conservative 90; Mismatches 205; Indels 211; Gaps 26;

Qy 22 RHQSEQRRELPKIVETESTSMDITIGSQKQPOFLKSIDELAFAFVAVETFKRQFDD--- 78
Db 1134 RNKAEKQKRDGEBLEALKTELEDTLDSTATQOELRAKE-----QEVTVLKKALDEETR 1188
Qy 79 -----LQKHIESIENAIKSLGNGVVLAAARNNHQPMSPRNNSVETTVTVS 129
Db 1189 SHEAQVQEMRQKHAQAVEE-LTEQLEQ--FKRAKANLDKNKOTLEKENADLAGELRV-LG 1244
Qy 130 QPSQEIIVPETSNTKEGGRMCMLM--CSKGLRKYYIANISDQA-KLMBEIPS----- 177
Db 1245 QAKQEV--EHKKKKLEAQVQELQSKCSDGER--ARAELNDKVHKLQNEVESVTGMLNEAE 1300
Qy 178 --ALKLAKEPAKF-----VLDCIGKFYLOGRRAFTK-----ESPM 210
Db 1301 GKAIKLAQVSLSSQLQDTQELLQOBTQKLVSTKRLQLEBERNSLODQDDEMEAKQ 1360
Qy 211 SSARQVSLILLESFLMPDRGK-----GKVKIESWIKD---EAETAAVAMR 253
Db 1361 NLERHISTLNQ-----LSDSKKLQDFASTVEALBEGKFKFQKEIENLTQOYEEKAAAYD 1416
Qy 254 KRLMTEGGL-----AAAEKMDARGLLLVACFGVPSNF----- 286
Db 1417 KLETKNLQOELDDLVLDNQQLVSNLEKKQKQKFDQLLAEEKNISKYADDERDRAEA 1476
Qy 287 ----RSTDLILDILIR-MGSGNEIAGALKR-SQFLVPVMSGIVESSIKRGWHIEALE----- 335
Db 1477 EAREKETKALSARALEEALAEKEELERTNKMKAEMEDLVSSKDDVGNVHELEKSKEA 1536
Qy 336 -----MVYTFGMEKDFSA-----LVLTSLFKMSKESFER----- 365
Db 1537 LETQMEEMKQLEBELELQATEDAKLRLEVNMQALKGQFERDLQARDQNEEKRLQOR 1596
Qy 366 -----AKRK-----AQSPFLAFKEAATKQALVLSVMQ- 392
Db 1597 QLHEYETELEDERKORALAAAAKKKLEGLDKLELQADSAIKGREBAIKQLRKLOQMKD 1656
Qy 393 -----CMETHKLDPAKELPGWQIKQEIQVLSLEKDTLQLDKEMEKEKARS 434

US-09-927-597-2
; Sequence 2, Application US/09927597
; Publication No. US20030032018A1
; GENERAL INFORMATION:
; APPLICANT: Malik, Fady
; APPLICANT: Beraud, Christophe
; APPLICANT: Freedman, Richard
; APPLICANT: Craven, Andrew
; APPLICANT: Sakowicz, Roman
; APPLICANT: Hartman, James
; TITLE OF INVENTION: Human smooth muscle myosin heavy chain
; FILE REFERENCE: CITOP018
; CURRENT APPLICATION NUMBER: US/09/927,597
; CURRENT FILING DATE: 2001-08-10
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 1945
; TYPE: PRT
; ORGANISM: Human
US-09-927-597-2

Query Match 3.7%; Score 114.5; DB 11; Length 1945;
Best Local Similarity 19.7%; Pred. No. 2;
Matches 124; Conservative 90; Mismatches 205; Indels 211; Gaps 26;

Qy 22 RHQSEQRRELPKIVETESTSMDITIGSQKQPOFLKSIDELAFAFVAVETFKRQFDD--- 78
Db 1141 RNKAEKQKRDGEBLEALKTELEDTLDSTATQOELRAKE-----QEVTVLKKALDEETR 1195
Qy 79 -----LQKHIESIENAIKSLGNGVVLAAARNNHQPMSPRNNSVETTVTVS 129
Db 1196 SHEAQVQEMRQKHAQAVEE-LTEQLEQ--FKRAKANLDKNKOTLEKENADLAGELRV-LG 1251
Qy 130 QPSQEIIVPETSNTKEGGRMCMLM--CSKGLRKYYIANISDQA-KLMBEIPS----- 177
Db 1252 QAKQEV--EHKKKKLEAQVQELQSKCSDGER--ARAELNDKVHKLQNEVESVTGMLNEAE 1307
Qy 178 --ALKLAKEPAKF-----VLDCIGKFYLOGRRAFTK-----ESPM 210
Db 1308 GKAIKLAQVSLSSQLQDTQELLQOBTQKLVSTKRLQLEBERNSLODQDDEMEAKQ 1367
Qy 211 SSARQVSLILLESFLMPDRGK-----GKVKIESWIKD---EAETAAVAMR 253
Db 1368 NLERHISTLNQ-----LSDSKKLQDFASTVEALBEGKFKFQKEIENLTQOYEEKAAAYD 1423
Qy 254 KRLMTEGGL-----AAAEKMDARGLLLVACFGVPSNF----- 286
Db 1424 KLETKNLQOELDDLVLDNQQLVSNLEKKQKQKFDQLLAEEKNISKYADDERDRAEA 1483
Qy 287 ----RSTDLILDILIR-MGSGNEIAGALKR-SQFLVPVMSGIVESSIKRGWHIEALE----- 335
Db 1484 EAREKETKALSARALEEALAEKEELERTNKMKAEMEDLVSSKDDVGNVHELEKSKEA 1543
Qy 336 -----MVYTFGMEKDFSA-----LVLTSLFKMSKESFER----- 365
Db 1544 LETQMEEMKQLEBELELQATEDAKLRLEVNMQALKGQFERDLQARDQNEEKRLQOR 1603
Qy 366 -----AKRK-----AQSPFLAFKEAATKQALVLSVMQ- 392
Db 1604 QLHEYETELEDERKORALAAAAKKKLEGLDKLELQADSAIKGREBAIKQLRKLOQMKD 1663
Qy 393 -----CMETHKLDPAKELPGWQIKQEIQVLSLEKDTLQLDKEMEKEKARS 434

US-09-927-597-2
; Sequence 2, Application US/09927597
; Publication No. US20030032018A1
; GENERAL INFORMATION:
; APPLICANT: Malik, Fady
; APPLICANT: Beraud, Christophe
; APPLICANT: Freedman, Richard
; APPLICANT: Craven, Andrew
; APPLICANT: Sakowicz, Roman
; APPLICANT: Hartman, James
; TITLE OF INVENTION: Human smooth muscle myosin heavy chain
; FILE REFERENCE: CITOP018
; CURRENT APPLICATION NUMBER: US/09/927,597
; CURRENT FILING DATE: 2001-08-10
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 1945
; TYPE: PRT
; ORGANISM: Human
US-09-927-597-2

Query Match 3.7%; Score 114.5; DB 11; Length 1945;
Best Local Similarity 19.7%; Pred. No. 2;
Matches 124; Conservative 90; Mismatches 205; Indels 211; Gaps 26;

Qy 22 RHQSEQRRELPKIVETESTSMDITIGSQKQPOFLKSIDELAFAFVAVETFKRQFDD--- 78
Db 1141 RNKAEKQKRDGEBLEALKTELEDTLDSTATQOELRAKE-----QEVTVLKKALDEETR 1195
Qy 79 -----LQKHIESIENAIKSLGNGVVLAAARNNHQPMSPRNNSVETTVTVS 129
Db 1196 SHEAQVQEMRQKHAQAVEE-LTEQLEQ--FKRAKANLDKNKOTLEKENADLAGELRV-LG 1251
Qy 130 QPSQEIIVPETSNTKEGGRMCMLM--CSKGLRKYYIANISDQA-KLMBEIPS----- 177
Db 1252 QAKQEV--EHKKKKLEAQVQELQSKCSDGER--ARAELNDKVHKLQNEVESVTGMLNEAE 1307
Qy 178 --ALKLAKEPAKF-----VLDCIGKFYLOGRRAFTK-----ESPM 210
Db 1308 GKAIKLAQVSLSSQLQDTQELLQOBTQKLVSTKRLQLEBERNSLODQDDEMEAKQ 1367
Qy 211 SSARQVSLILLESFLMPDRGK-----GKVKIESWIKD---EAETAAVAMR 253
Db 1368 NLERHISTLNQ-----LSDSKKLQDFASTVEALBEGKFKFQKEIENLTQOYEEKAAAYD 1423
Qy 254 KRLMTEGGL-----AAAEKMDARGLLLVACFGVPSNF----- 286
Db 1424 KLETKNLQOELDDLVLDNQQLVSNLEKKQKQKFDQLLAEEKNISKYADDERDRAEA 1483
Qy 287 ----RSTDLILDILIR-MGSGNEIAGALKR-SQFLVPVMSGIVESSIKRGWHIEALE----- 335
Db 1484 EAREKETKALSARALEEALAEKEELERTNKMKAEMEDLVSSKDDVGNVHELEKSKEA 1543
Qy 336 -----MVYTFGMEKDFSA-----LVLTSLFKMSKESFER----- 365
Db 1544 LETQMEEMKQLEBELELQATEDAKLRLEVNMQALKGQFERDLQARDQNEEKRLQOR 1603
Qy 366 -----AKRK-----AQSPFLAFKEAATKQALVLSVMQ- 392
Db 1604 QLHEYETELEDERKORALAAAAKKKLEGLDKLELQADSAIKGREBAIKQLRKLOQMKD 1663
Qy 393 -----CMETHKLDPAKELPGWQIKQEIQVLSLEKDTLQLDKEMEKEKARS 434
```

Db 1664 FORELEDARSRDEIFATAKENEKAKSLEADLMQLOEDLAAMERARKQADLEKEELAE 1723

QY 435 ISLMEEAALAKEMVNOQIKR---PRLSPE 461

Db 1724 LA---SSLSGRNALQDEKRRLEARIAQLE 1749

Search completed: August 14, 2003, 10:49:26
Job time : 62 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 14, 2003, 10:34:59 ; Search time 42 Seconds
(without alignments)
1394.446 Million cell updates/sec

Title: US-09-890-475-1

Perfect score: 3104

Sequence: 1 MSNYPPTVAAGQTTTANPLL.....RYLGLSNQRSPRNSLIDPK 609

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 9616862 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 75:**

1: PIR1:**

2: PIR2:**

3: PIR3:**

4: PIR4:**

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Query %	Length	DB ID	Description
1	2653.5	85.5	578	2	T01229	hypothetical prote
2	2295	9.5	473	2	A86442	hypothetical prote
3	286.5	9.2	862	2	T01798	hypothetical prote
4	234.5	7.6	507	2	C71412	probable hydroxypr
5	142	4.6	725	1	JC5016	hyaluronan recepto
6	139.5	4.5	1201	2	T08603	kinesin-related pr
7	136	4.4	1722	1	I78879	retinoblastoma bin
8	134	4.3	841	2	A86188	hypothetical prote
9	134	4.3	2116	2	A26555	myosin heavy chain
10	132	4.3	5327	2	T13564	microtubule-associ
11	131	4.2	993	2	S49461	synaptonemal compl
12	129	4.2	946	2	S28061	SCP1 protein - rat
13	127.5	4.1	2897	2	A48666	cell proliferation
14	127.5	4.1	3256	2	A48666	cell proliferation
15	126.5	4.1	602	2	T13988	aryl hydrocarbon r
16	126.5	4.1	1004	2	B69483	hypothetical prote
17	126.5	4.1	1920	2	A53188	pericentrin - mous
18	126	4.1	694	2	T25685	hypothetical prote
19	125	4.0	1128	2	A49960	bud emergence prot
20	125	4.0	1974	2	T16703	hypothetical prote
21	124.5	4.0	1127	2	T32404	hypothetical prote
22	124.5	4.0	1538	2	T29095	cardiac muscle fac
23	124	4.0	3488	2	T34418	hypothetical prote
24	122.5	3.9	650	2	T00081	hypothetical prote
25	122.5	3.9	1034	2	T32297	hypothetical prote
26	122.5	3.9	2285	2	T12796	probable transglyc
27	122	3.9	1325	2	T42722	male-enhanced anti
28	121.5	3.9	279	2	D71453	hypothetical prote
29	121.5	3.9	3187	2	JC5837	364K Golgi complex

serine/threonine p
myosin-like protei
protein F23M19.10
hypothetical prote
hypothetical ABC t
protein C4IG7.5 [i
protein K09H11.3 [i
serine/threonine-s
membrane associat
still life protein
still life protein
hypothetical prote
kinesin-related pr
early endosome ant
hypothetical prote
hypothetical prote

30 120.5 3.9 1001 2 T17365
31 120.5 3.9 1875 2 S38173
32 120 3.9 619 2 E86466
33 118.5 3.8 627 2 A52714
34 119.5 3.8 637 2 E97496
35 119.5 3.8 650 2 G87883
36 119.5 3.8 1085 2 G89056
37 119.5 3.8 1388 2 S70633
38 119 3.8 680 2 A97331
39 118.5 3.8 2044 2 T13704
40 118.5 3.8 2064 2 T13707
41 118 3.8 991 2 H86168
42 117.5 3.8 1060 1 A40264
43 117.5 3.8 1410 1 A57013
44 116.5 3.8 700 2 C86296
45 116.5 3.8 1181 2 T01799

ALIGNMENTS

RESULT 1

T01229
hypothetical protein F6N23.25 - Arabidopsis thaliana
C1Species: Arabidopsis thaliana (mouse-ear cress)
C1Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 22-Oct-1999
C1Accession: T01229
R1Geisel, C.
submitted to the EMBL Data Library, April 1998
A1Reference: The sequence of A. thaliana F6N23.
A1Accession: T01229
A1Status: translated from GB/EMBL/DBJ
A1Molecule type: DNA
A1Residues: 1-578 <GEI>
A1Cross-references: EMBL:AF058919; NID:G3047100; PID:G3047115; GSPDB:GNC0063; ATSP:F6N2.
C1Genetics:
A1Gene: ATSP:F6N23.25
A1Map position: 5
A1Introns: 313/3; 530/1

Query Match 85.5%; Score 2653.5; DB 2; Length 578;
Best Local Similarity 83.6%; Pred. No. 2.8e-164;
Matches 531; Conservative 3; Mismatches 8; Indels 93; Gaps 2;

QY 1 MSNYPPTVAAGQTTTANPLLQHQSEQRRLPKIVETESTSMDITIGQSKQPOFLKSID 60
DB 1 MSNYPPTVAAGQTTTANPLLQHQSEQRRLPKIVETESTSMDITIGQSKQPOFLKSID 60
QY 61 ELAASFVAVEIFKRFQDDLOKHIESIENAIKLSNGVWLAARNNNFHQPMSPRRNV 120
DB 61 ELAASFVAVEIFKRFQDDLOKHIESIENAIKLSNGVWLAARNNNFHQPMSPRRNV 120
QY 121 SVETTVTSQSQEIVPETSNNKPEGRMCELMCKGURKIYANISDAKMEIIPSAK 180
DB 121 SVETTVTSQSQEIVPETSNNKPEGRMCELMCKGURKIYANISDAKMEIIPSAK 180
QY 181 LAKEPAKFVLDICIGKFLYQGRRAFTKESPMSSARQVSLILLESFLLMPDRGKGVKIESW 240
DB 181 LAKEPAKFVLDICIGKFLYQGRRAFTKESPMSSARQVSLILLESFLLMPDRGKGVKIESW 240
QY 241 IKDEAETAAVAKRLMTTEGGGLAAAKDAKGLLLLVACFGVPSNFRSTDLDIRMSG 300
DB 241 IKDEAETAAVAKRLMTTEGGGLAAAKDAKGLLLLVACFGVPSNFRSTDLDIRMSG 300
QY 301 NIAGALKRQSLVPMWVGIVESSIKRGMHTEALEMVTYTFGMEDKFSAALVLTSLKMSK 360
DB 301 NIAGALKRQSLVPMWVGIVESSIKRGMHTEALEMVTYTFGMEDKFSAALVLTSLKMSK 360
QY 361 ESFERAKRKAQSPFAFKAATKQALVSSVQCMETHKLDPAKELPGWIKQIVSLEKD 420
DB 361 ESFERAKRKAQSPFAFKAATKQALVSSVQCMETHKLDPAKELPGWIKQIVSLEKD 420
314 -----KEAATKQALVSSVQCMETHKLDPAKELPGWIKQIVSLEKD 357

QY 421 TLQDKEMEERKARSLIMEEAAIAKRYNNOQIKRPLSPNEMPPVTSSSYSPYDRDSEP 480
DB 358 TLQDKEMEERKARSLIMEEAAIAKRYNNOQIKRPLSPNEMPPVTSSSYSPYDRDSEP 417
QY 481 SORDDDDQDEISALVSSYLGPGTSPPHRSRRSPETWVPLPHGGIGRSYVAYEHLAPNSYSP 540
DB 418 SORDDDDQDEISALVSSYLGPGTSPPHRSRRSPETWVPLPHGGIGRSYVAYEHLAPNSYSP 477
QY 541 GHGHLRHQYSPSLVHGQORHPLQSPYPIHGQQQLPYGIQRYVPHSPSEERY-----591
DB 478 GHGHLRHQYSPSLVHGQORHPLQSPYPIHGQQQLPYGIQRYVPHSPSEERY-----537
QY 592 -----LGLSNQSPRSNNS 605
DB 538 MNVMSAFLSLRXWNSPKIQLELDNQNPKQKQS 572
RESULT 2
A86442
Hypothetical protein F5M6.26 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Dec-2001
C:Accession: A86442
R:Theologos, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Frazer, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; PMID:21016719; PMID:11130712
A:Accession: A86442
A:Status: preliminary
A:Map position: 1-473 <STO>
A:Residues: 1-473
A:Molecule type: DNA
A:Cross-references: GB:AE005172; NID:g11136721; PIDN:AAG31302.1; GSPDB:GN00141
C:Genetics:
A:Map position: 1
Query Match 9.5%; Score 295; DB 2; Length 473;
Best Local Similarity 22.5%; Pred. No. 1.6e-11;
Matches 124; Conservative 109; Mismatches 219; Indels 98; Gaps 21;
QY 39 ESTMDITIGSKQFQFLKSIDEIAAP-SVAVFETKQFDDLOKHIESIENAIKSLN 97
DB 5 ESTAASINQIDKQKLLKAFDDLOAHRSLLSPSNLSWSEIDSHFSLSQSLFNRLQS- 63
QY 98 GVVLAARNHFQPLMSPRNNVSVETTVTSQSQEIVPETSNNKPGGRCMLCKSL 157
DB 64 -----AVTSN-----SGNIETPAVTTETPVL-----WPLRFECKNDCKGL 102
QY 158 RKYIYANISDQAKLMEIPSAKLAKEPAKFLVDCIGKFLYQGRRAFTKSPMSSAR--- 214
DB 103 GNYMTENSRKLSINEELPAIRCSNPAPLVDAI-----EGSYHCSPPSSSSSARAID 157
QY 215 --QVSLILLESFLMPDRCKGKVKIESWKDEATAVAVKRLMTWEGGLAAEKMDARG 272
DB 158 VKRIFVILLLEALIEI-----NANUTNDRERARTIADWKPNI-----GNKPSEALG 204
QY 273 LLLLACFGVPSNFRSTDLDDLRMSGNEIAGAL-KRSQFLVPMVSGIVESSIKRGHMI 331
DB 205 FLHLVAAPFLGSLFSTEEICDYIFLSKYQATTICKIGLDRNRIGVYVQKFLDTGRLL 264
QY 332 BALEWVYTFGMDKSAALVLTSLFKMSKESFERAKRAQSPAPK-BAATKQALVLSV 390
DB 265 VAIRFIYENVMGSEFEPVSLTSLKREAKRVCAEGNYSLVKQVNEATKELSLARV 324
QY 391 MOCMETHKLDPAKLPGMQIKQIVLSKDTLQDKEMEERKARSLIMEEAAIAKRYNNO 450
DB 325 IKVVKEKNIES-----EFMEE--KLEECVKELEDQAKQKRA-TKFNSPANPQOPQEQ 374

QY 451 QI--KRPRL---SPME---MPVTSSSYSPYDRDSEPSPQSDDDQDEISALVSSYLGPS 501
DB 375 KVNCKPRVANGSMSEYNTIPPLRPQQPPPL-----LPTSQILQVNPYGLLSAIL-PG 428
QY 502 TSPFHRSSRSPETWVPLPHGGIGRSYVAYEHLAPNSYSPGHGHLRHQYSPSLVHGQORHP 561
DB 429 VAVPYGNPRALFGSVAP---ASRPVYVQ-----QTGYG-----MPP 463
QY 562 LOYSPPIHGQ 571
DB 464 PQYRPPYYPQ 473
RESULT 3
T01798
Hypothetical protein A.TM021B04.8 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 26-Feb-1999 #sequence_revision 26-Feb-1999 #text_change 22-Oct-1999
C:Accession: T01798
R:Dante, M.; Wamsley, P.; Gibson, A.
submitted to the EMBL Data Library, June 1997
A:Description: The sequence of A. thaliana TM021B04.
A:Reference number: Z14440
A:Accession: T01798
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-862 <DAN>
A:Cross-references: EMBL:AF007271; NID:g2191181; PID:g2191193; GSPDB:GN00063; ATSP:A_TMO
C:Genetics:
A:Gene: ATSP:A.TM021B04.8
A:Map position: 5
A:Indels: 220/1; 280/3; 379/1; 418/1; 615/1; 674/3
Query Match 9.2%; Score 286.5; DB 2; Length 862;
Best Local Similarity 22.1%; Pred. No. 1.3e-10;
Matches 132; Conservative 119; Mismatches 231; Indels 115; Gaps 21;
QY 18 PLLQHQSEQRRELKIVETESTMDITIGSKQFQFLKSIDEIAAPSVAVETKQFQD 77
DB 355 PYEQKHVKKR-----PLTETPTAPSNQTLVCFNMKHELTSVNULSIF-VFGETLRRHF- 407
QY 78 DLQKHIESIENAIKSLNGLSVVLAARNNN-----PHQPLMSPRNNVSVETTVTSQ 130
DB 408 -----NQNIFLLFKLKQPEV-----VCVPTGKQVKE 434
QY 131 PSQIVPETSNNKPGGRCMLCKSG-----LRKYIYANISDQAKLMEIPSAKLAKE 184
DB 435 SGADHQPDITATPHSGTETKLNILSGSIKADMURELVE---KQPLKESDLSNALKCTPD 491
QY 185 PAKFVLDICIGKFLYQGRRAFTKSPMSSARQVSLILLESFLMPDRCKGKVKIESWKDE 244
DB 492 PAKFLDTSMALCPTNTEGGYEFKMLITSASCSSLNQLKLLP-----KIGHVPKGD 544
QY 245 AETAAVAVKRLMTWEGGLAAEKMDARGLLLVACGVPSNFRSTDLDDLRMSGNEIA 304
DB 545 AKKLAVYWKDKI-----AKSRDQLEVICFQIGIFVSEFKADDDLLGLDLSNYQVTS 600
QY 305 GALKESQFLVPMVSGIVESSIKRGHIEALEWYTFGMDKFS-AALVLTSLFKMSKESF 363
DB 601 PDLQCFGLDDAIFGFTQNLIKTGHRIKADYIYSGMVHRFQPVSAIINDSLRITKESA 660
QY 364 BRAKRAQSPAPKFAA-TKQALVSSVMOCMETHKLDPAKELPFGWQIKQIVLSKDTIL 422
DB 661 EKSRYEAKNSETTQVAAIDRQVRAALAAIKCISCHKESEFQL--GDLBEQIKSLK--L 716
QY 423 QLDKEMEERKARSLIMEEAAIAKRYNNOQIKRPLSPNEMPPVTSSSYSPYDRDSEP 482
DB 717 RNTNSGSGSGSASKPDSTIKQ---SQAKPTVA--EVAPTS-----NIPLE 761
QY 483 RDDDDQDEISALVSSYLGPGTSPPHRSRRSPETWVPLPHGGIGRSYVAYEHLA--FN 536
DB 762 PSTEAASSA-----SKPFSKKNKR-----GKKRSMGNNQSSGSHASHTS 802

QY 537 SVSPGCHRLHROVSPSLVGHQHPLOVSPPIHQQQ--LPYGIQVYRHSPSEERY 591
DB 803 NHYPSHDYSNQLRTPVDNYDRGTGTFPNDYNNQWQEGPQFYHLYQPLDPY 859

RESULT 4
C71412
hyaluronan receptor - human
C:Species: Arabidopsis thaliana (mouse-ear cress)
A:Variety: Columbia
C:Date: 03-Aug-1998 #sequence_revision 03-Aug-1998 #text_change 05-Dec-1998
R:Sevan, M.; Bancroft, I.; Bent, E.; Love, K.; Goodman, H.; Dean, C.; Bergkamp, R.; Dirk
P.; Wedler, H.; Wedler, B.; Wambutt, R.; Weitzenecker, T.; Pohl, T.M.; Terry, N.; Giel
avanagh, T.; Hempel, S.; Kotter, P.; Entian, K.D.; Rieger, M.; Schaeffer, M.; Funk, B.
Nature 391, 485-488, 1998
A:Authors: Mueller-Auer, S.; Silvey, M.; James, R.; Montfort, A.; Pons, A.; Puigdomenech
erhoft, A.; Moore, T.; Jones, J.D.G.; Eneva, T.; Palme, K.; Benes, V.; Reckman, S.; Ana
C.; Chalvatzis, N.
A:Title: Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of Arabidopsis thal
A:Reference number: A71400; MUID:98121113; PMID:9461215
A:Accession: C71412
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-507 <BEV>
A:Cross-references: GB:297337; NID:g2244829; PID:e326836; PID:g2244847
C:Genetics:
A:Map position: 4C09-4G3845

Query Match 7.6%; Score 234.5; DB 2; Length 507;
Best Local Similarity 19.2%; Pred. No. 1.4e-07;
Matches 117; Conservative 90; Mismatches 214; Indels 187; Gaps 21;

QY 49 QSKQPOFLKSIDELAFAFVAVETPKRQ-----FDDLOKHIESIENAIIDSKLSNG 98
DB 12 KSSQSFEE-----FQQASLMTSCNLLMKELSEHTSMEQNLMKSEALR 57

QY 99 VVLAARNNHFHOPMLSPNNVSVETTVTS-----QPSQIVPETSNDP 143
DB 58 QMIETLDNQTOSIIEELKREVTIDHSVEIAEGKVEERVRAALDSLEKARDCGEDTGEV 117

QY 144 EGG-----RMCELMCKSLGRLKIYANISDQAKLMEIPSAKLAKPAKFVLDICGK 195
DB 118 DGDGGLSALKSLCKLDKARGFWGFIARKLENLRSQIPVALVDCVDPKLVLEAVSE 177

QY 196 FYLQGRRAFTKESPMSSARQVSLILLESF---LLMPDRGKGVKIESWIKDEATAVAV 252
DB 178 VFPVKRG-GEKVSNDFGWACVWVLESIPVMVDPVMGKSLVTPSVYKAKEIAETW 236

QY 253 RKRLMTEGGLAAEKMDARGLLLLVACFGVPSNFRSTDLDLIRMSGNEIAGALKRSQF 312
DB 237 KASLEERGGIENKTPDVHTFLOHVTFGIVKK---DDLALYR---KLVVGSARWQOM 280

QY 313 -----LVPVMSGIVGESIKRGHIEALEVMVYTFGMEDEKFSALVLTSPKMSKESFE 364
DB 289 PKLAVSVGLQDQPMTEELIIRGQQLDAVHTFVEGLVHLFPVPLLKAYLRDAKATA 348

QY 365 -----RAKKAQSPFLAFKAAATKQALVLSVWQCMETHKLDPAKELPCHWIKSQI 414
DB 349 LIITDSDNNSGRSAGSVSSALAY-----QCMLEKTKTKRKA-----V 387

QY 415 VSLKEDTLQDKEMEEKARSLMEBAALAKRMYNQOIKRPLSPMPPVTSSTSPY 474
DB 388 IPANKRT-----RASYSG-----PMPPAKAGRIT--- 411

QY 475 RDRSFPSQRDDQDEISALVSVYLGSTSPPHRSRSPSPYPLPHGGLGRVYAYEHLA 534
DB 412 -----NAVYSSFPPTPTFIRSQSHSPQYGV-----AYTTSP 444

QY 535 PNV---SPGHGHLHROVSPSLVGH--QRHPLQY-----SPPI-HGQQOL 574
DB 445 PTIYNRSPPY-----QYSPFAVHGSYQTSFYSYATGYTCVSPVAAAPPVPHPHPH 498

QY 575 PYGIQVRY 582
DB 499 HHHQHAY 506

RESULT 5
JC5016
hyaluronan receptor - human
C:Species: Homo sapiens (man)
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 29-Sep-1999
R:Wang, C.; Entwistle, J.; Hou, G.; Li, Q.; Turley, E.A.
Gene 174, 299-306, 1996
A:Title: The characterization of a human RHAMM cDNA: Conservation of the hyaluronan-bin
A:Reference number: JC5016; MUID:97045829; PMID:8890751
A:Contents: breast
A:Accession: JC5016
A:Molecule type: mRNA
A:Residues: 1-725 <WAN>
A:Cross-references: GB:U29343
A:Note: it is uncertain whether Met-1 or Met-196 is the initiator
C:Comment: This receptor regulates focal adhesion turnover, and regulates human breast
C:Genetics:
A:Gene: GDB:HMWR; RHAMM
A:Cross-references: GDB:683209; OMIM:600936
A:Map position: 5q33.2-5qter
C:Superfamily: hyaluronan receptor

Query Match 4.6%; Score 142; DB 1; Length 725;
Best Local Similarity 20.0%; Pred. No. 0.24;
Matches 112; Conservative 91; Mismatches 220; Indels 136; Gaps 21;

QY 34 KIVETESTMDITTCQSKQPOFLKSIDELAFAFVAVETPKRQFDDLOKHIESIENAIIDSKLS 92
DB 216 KLVSTKEKID---EKSETKLEYIEISCASQVQEKYKLDIAQLEENLKEKNDEILSL 272

QY 93 --KLASNGVLA-----ARNNFHQPMLSPNNVSVETTVTSQ 130
DB 273 KOSLEENIVLSKQVEDLNVKCOLLEKEKEDHVNREHNENLNAEQNL----- 322

QY 131 PQSEIVPETSNDPKNPEGRVCMELCKSLGRLKIYANISDQAKL---MEEIPSAKLAKPAK 187
DB 323 -KQKFILE--QSEHEKLOQKELQIDSLLOQKELSSSLHQLCSPEEMVKEKNLFEELK 380

QY 188 FVLDCIGKTYL---QGRRAFTKESPMSSARQVSLILLESFLLMPDRGKGVKIESWIKDE 244
DB 381 QTLDELKLOQKEQAEERLVQLEEEAKSRAEELKLEEL---KGK-----EAE 427

QY 245 AETAAVANERKRLMTEGGLAAEKMDARGLLLLVACFGVPSNFRS----- 288
DB 428 LEKSSAA-----HTQATLLQEKYDS---MVQSLQEDVTAQFESYKALTASEIEDLKLEN 478

QY 289 TOLLDIRMSGN-----EIAGALKRSQFLVPMVSGIVGESIKRGHIEALEVMVYTFGM 342
DB 479 SSIQEKAAKAGKNABEDVQHIILATESSNQEVYRMLDLQTSALKETEIKETVTSFLQKI 538

QY 343 EDKFSALVLTSLFKMSKESPERAKRKAQSPFLAFKAAATKQALVLSVWQCMETHKLDPA 402
DB 539 TD-----LQNLQKQBEDFRKQLEDEEGRKAENKTAEI-----T 574

QY 403 KEIPLGHQ-KEQIVSLKED-TLQDK-EMEEKARSLMEBAALAKRMYN----- 449
DB 575 EEINKVRLLYEELYNKTKPFLQDLDAFEVEKQA-----LLNEHGAAQEOQLNKIRDSYAKLL 630

QY 450 --QOIKRPLSPMPPVTSSTSPYIRDRGPPSPQRDDQDEISALVSVYLG-----PST 502
DB 631 GHQNLQKIKHVVKLKDNESQKLSKRLCQAKKQSETKLQELNKLVLGIKHFDPSK 690

QY 503 SFPHRSRSPSPYPLPHG 521
DB 691 AFHESKENFALKTPLKEG 709

RESULT 6

T08603
 kinesin-related protein COS2 - fruit fly (*Drosophila melanogaster*)
 C:Species: *Drosophila melanogaster*
 C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 17-Nov-2000
 C:Accession: T08603
 R:Stinson, J.C.; Ho, K.S.; Suyama, K.; Scott, M.P.
 Cell 90, 235-245, 1997
 A:Title: Costal2, a novel kinesin-related protein in the Hedgehog signaling pathway.
 A:Reference number: Z16453; MUID:97386449; PMID:9244298
 A:Accession: T08603
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-1201 <SIS>
 A:Cross-references: EMBL:AF019250; NID:92330020; PIDN:AAB66813.1; PID:92330021
 C:Genetics:
 A:Gene: costal-2
 A:Cross-references: FlyBase:FBgn0000352
 A:Map position: 2
 C:Keywords: cytosol; microtubule binding; signal transduction

Query Match 4.5%; Score 139.5; DB 2; Length 1201;
 Best Local Similarity 19.7%; Pred. No. 0.74;
 Matches 132; Conservative 97; Mismatches 233; Indels 209; Gaps 29;

Db 18 PFLQRHQSEORRELKIVETESTMDITIGSQKQPFKLSIDELAFAVAVETFKRQFD 77
 498 PILEAEEDVNSEANSESNSENDNTDNESHRLDDKIESL-----MEEFKDT 551
 78 D-----LQKHTESTENAIKLES-----NG-----VVLAAARNNNFHQPMSPRRN 119
 552 ALILEKHAELSKPKAVMQQDRIEAQPPENGDDRKVSIGSRRSV-QPGASLSTAE 610
 120 VSEVTTVTVSQSQIIVPET-----SNKPEGGVMCELMCS-----KGLRKYIVA 163
 611 LAMLRNVAQQPPPIDESVVDPLESSSGEIGIQALAAATAPIEQOKLRKLV-A 669
 164 NISDAQKLMEEIPSAKLA-----KEPAKFVLD-----IGK 195
 670 EIEGKQRLREIETIQVKNIIAELVNSDTRSHAKQRFHKKALEACDRAKQLGK 729
 196 FYLQGR-----RAFTKESPM-----SSAQVSL-----ILESFLMPDRG 231
 730 ALVQGRGQSEIERWTIIGHLERLLEDLSMKHTAGESGQVKLQOSVGESRKQADDLQ 789
 232 KGVKIESWIKDEATAAVAKRLMTTEGLAAAE---KMDARGLLLVACFGVPSNFRS 288
 790 K-KLRKECKLRQMEAEVLKRESRETGKELVKAQGSPEQOQRLKAVQA-----RI 840
 289 TDLILRMGSGN-----ETAGALKRSQFLVPM 316
 841 TLNHLREKSNLBEQPGPQQETLRHEIRNLRTDRLLEECHELDRLKRDKVLTKQ 900
 317 VSGIVESSTKRGWHLEALEMYVTFGMEDKFAALVLTSLFKMSKESPERAKRKAQPLAF 376
 901 EE---RKLECEDEATEADAIEFPNEMITGHRSDTSD-RIQREKGEQMLMARNLST 956
 377 KBAATKQALVSSVMQC-----NETHKLDPAKLPQWQKEQIVSL-----EKOT 421
 957 EEMRTLLYKYFTKVIDLRDSRKLEQLVQLERPERDAWKEKRVLSNVAQARLEGNERA 1016
 422 LQDKMEMEERKARLSIM-----EEAALAKRMYNQIKRPRLSMPMPFPVTSSS---YSPI 473
 1017 VLLQRCHENK---LTLMLRHAEETISASSASYGERALAP---ACVAPPVQASSDFDYDH 1070
 474 YDRFPSPQRDDQDEIGALVSYLGPSTSPHRSRSPETMYPLPHGL-----523
 1071 YKGGGNFSK-----ALI-----KAPK---PMPTGSAIDKYKQEQR 1103
 524 -GRSVYAYEHL 533
 1104 SGRNIFAKFHV 1114

RESULT 7

178879
 retinoblastoma binding protein 2 - human
 C:Species: *Homo sapiens* (man)
 C:Date: 17-Mar-2000 #sequence_revision 17-Mar-2000 #text_change 17-Mar-2000
 C:Accession: 178879; S16954
 R:Pataky, A.R.; Helin, K.; Dembski, M.S.; Dyson, N.; Harlow, E.; Vuocolo, G.A.; Hanobik
 Oncogene 8, 3149-3156, 1993
 A:Title: Characterization of the retinoblastoma binding proteins RBP1 and RBP2.
 A:Reference number: 158383; MUID:94020841; PMID:8414517
 A:Accession: 178879
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-1722 <EAT>
 A:Cross-references: GB:S66431; NID:9435777; PIDN:AAB28544.1; PID:9435778
 R:Deleo-Jones, D.; Huang, P.S.; Jones, R.E.; Haskell, K.M.; Vuocolo, G.A.; Hanobik, M.G.
 Nature 352, 251-254, 1991
 A:Title: Cloning of cDNAs for cellular proteins that bind to the retinoblastoma gene pro
 A:Reference number: S16953; MUID:91312450; PMID:1857421
 A:Accession: S16954
 A:Status: not compared with conceptual translation
 A:Molecule type: mRNA
 A:Residues: 1102-1562, 'KKK' <DEF>
 C:Genetics:
 A:Gene: GDB:RBP2
 A:Cross-references: GDB:119548; OMIM:180280
 A:Map position: 3q21-3qter
 C:Superfamily: human retinoblastoma binding protein 2

Query Match 4.4%; Score 136; DB 1; Length 1722;
 Best Local Similarity 19.1%; Pred. No. 2.1;
 Matches 146; Conservative 120; Mismatches 254; Indels 244; Gaps 38;

QY 14 TTANFLQRHQSEORRELKIVETESTMDITIGSQKQPFKLSIDELAFAVAVETPK 73
 Db 796 SVAQLLSK---KQKHQSP---DSGRTRTKLIVELK-AFVQQLFSLPCVISQARQVK 847
 QY 74 RQFDLQKHIESIENAI-----DSKLESNVVLAARNNNFHQPMSPRRNVSV----- 122
 848 NLLDVEEFHERAQEAAMMDTPDSSKLQ-----MLIDMGSSLYVELPELRLKQLQQAQW 903
 QY 123 --ETVTVSQPSQEIIVETSNKPEGG-----RMCEL-----MCS 154
 904 LDEVRLTDPQOQVTLDMVKKLDSGVGLAPHAVKAEKMAELQELLTVSERWEEKAKVCL 963
 QY 155 KGLRKYIYANISDAQKLMEEIPSAKLAKEPAKFVLDCTGKFYLOGRRAFTK--ESPMS 212
 964 QARPHSVASLESIVNEAKNIPAPL-----PNVLSL-----KEALQKAREWTAKVEAIQSG 1014
 QY 213 ARQVSLILLESFLMPDRGKKG---VKIESWIKDEATAAV-AWRKRLMTTEGLAAAEKW 268
 1015 SNAYLBQLESL-----SAKRPIPVRLREALPOVESQVAAARWRRT-----GRTFLKN 1065
 269 DARGLLLVACFGVPSNFRSTDLDIRMSGSNEIAGALKRSQFLVPMVSGIVESSIKRG 328
 1066 SSHLLQVLS-----PRTD---IGVVGSG---GKNRRK---VKELIEKEKED 1104
 QY 329 MHIALEMYVTFGMEDKFAALVLTSLFKMSKESPERAKRKAQSPFAKFAATKQLAVLS 388
 1105 LDLEPLSDLEE-GLEETRDYAMVAVFK-----BRQKEIEAMSHSLRAANLAKMTWYD 1156
 QY 389 SVMQ-----CMETHKLDPAKELPGWQIKE-----412
 1157 RIEEVKFCICRTASGFMQLQCELCQWDFHNSCVPLPKSSQKKGSSQWAKEVFKLCLPCM 1216
 413 -----QIVSLEKDTQLDK-----EMEEKAR-----SLSLME 440
 1217 RSRPRLETILSLVSLQKLPVLPSEALQCLITERAMSQDRARQALATDELSSALAKL 1276
 QY 441 AALAKRMYNQ-----QIKRPLSP---MMPPVTSSYSPI-----473

Db 1277 SVLSQRMWQAAREKTEKIISAELOKAAANPDLOHLPSPFOOSAFNRVSVSSSPQTM 1336
QY 474 -YRDRSFPQRD-----DDQDEISALVSSYLGPS-----TSFPHRRSRSPYVPLPHG 521
Db 1337 DYVDEETDSDEIRYGYDMKDTASVKSSSLEPNLFCDBEIPKBEVVTHMTAPS- 1395
QY 522 GLGRSVVAEHL---APNSYSP---GHGHLRQY---SPSLVHGQRHP-LOYSPPTHQ 571
Db 1396 -----FCAEHAYSSASKSCSVFFGKSGSTPRKQPSPLVRSLEPPVLELSPGAKA 1449
QY 572 --QQLPYG-----IQRVYR-----HSPSEERYLGLSNQRS 599
Db 1450 LSELMMVGLLEVSLEDTCHINRIILQATHPPSEDRFLHIMEDDS 1493
RESULT 8
A86188
hypothetical protein [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C;Accession: A86188
R;Theologos, A.; Ecker, J.R.; Palm, C.J.; Pederspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.; Shann, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Walzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719; PMID:11130712
A;Accession: A86188
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-841 <STO>
A;Cross-references: GB:AE005172; NID:G2388564; PIDN:AAB71445.1; GSPDB:GN00141
C;Genetics:
A;Map position: 1
Query Match 4.3%; Score 134; DB 2; Length 841;
Best Local Similarity 20.8%; Pred. No. 1;
Matches 118; Conservative 74; Mismatches 196; Indels 180; Gaps 21;
QY 20 LQHQSEQRRLPKIVETESTSMDITIQ-----SKQPFLLKSIDEL-----AAPS 66
Db 132 LQK-QSASELEKXISDERYKTDALLSQALSONSLVLEKLSLEELSEKVSLEKLSALI 190
QY 67 VAVETKQFDDLOKHLESTENAIKSLKESNGVLAARNNNHQPW----- 112
Db 191 VAEERGKSSIQMOYOEKV-----SKLESSLNQSARNSELEEDRLALOGAHEDIG 245
QY 113 -LSPPRN-----NVSVETTVTSQPSQEIYVETSN 141
Db 246 NVSTKRSVELQGLFQTSQKLEKAEKLDLEAIQVNSSLEATLSVAMEKRDLSNLN 305
QY 142 KPEGRCMELCMCKGLRKIYANISQA-----KLMEIPALKLA 182
Db 306 -----AVMEKLSSEERLEKQAREIDEATRSIEALHKKSELKVKQTMEDFSRDTA 360
QY 183 KEPAKFVLDICIGFYK-QGRRAFTKESPMSSARQVSLILLESFLLMPORGKGVKI---E 238
Db 361 KSLTERSKDLKEIRVYEGKLAECAGSLSLQBELQSSAENELLADTNQLKIKIQELE 420
QY 239 SWIKDEAETAARVWRKMLTEGGLAAEKMDARGLLLLVACFGVSNFRSTDLILLRMS 298
Db 421 GYLDSEXET-----ATKEL-----NQKDTAKDLITKL 448
QY 299 GSNEIAGALKESQPLVPMVSGIVESSIKRGWHTALEMVTTFGMEDKFGAALVLSFLKM 358
Db 449 KSHD--NVIEHKQVLEASGVADT---RKVEVE-----EALLKNTLES 488
QY 359 SKESFER-----AKKAQSPILAFKAATQKLAVLSSVQCWETHKLDPAKELPGWQIK 411

RESULT 9

A26655
myosin heavy chain [similarity] - slime mold (Dictyostellium discoideum)
N;Contains: myosin Atpase (EC 3.6.4.1)
C;Species: Dictyostellium discoideum
C;Date: 05-Oct-1988 #sequence_revision 05-Oct-1988 #text_change 19-Apr-2002
C;Accession: A26655; A24728; S00250
R;Warrick, H.M.; De Lozanne, A.; Leinwand, L.A.; Spudich, J.A.
Proc. Natl. Acad. Sci. U.S.A. 83, 9433-9437, 1986
A;Title: Conserved protein domains in a myosin heavy chain gene from Dictyostellium disc
A;Reference number: A26655; MUID:87092266; PMID:3540939
A;Accession: A26655
A;Molecule type: DNA
A;Residues: 1-2116 <WAG>
A;Cross-references: GB:M14628; GB:M11938; NID:G167834; PIDN:AAA33227.1; PID:G167835
R;DeLozanne, A.; Lewis, M.; Spudich, J.A.; Leinwand, L.A.
Proc. Natl. Acad. Sci. U.S.A. 82, 6807-6810, 1985
A;Reference number: A24728; MUID:86016788; PMID:3901008
A;Accession: A24728
A;Molecule type: mRNA
A;Residues: 2035-2116
R;Wagle, G.; Noegel, A.; Scheel, J.; Gerisch, G.
FEBS Lett. 227, 71-75, 1988
A;Title: Phosphorylation of threonine residues on cloned fragments of the Dictyostellium
A;Reference number: S00250; MUID:89112226; PMID:2828113
A;Accession: S00250
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1734-1893 <WAG>
C;Comment: The rod domain is highly periodic, containing a pattern of 7-residue repeats
C;Superfamily: myosin heavy chain; myosin motor domain homology
C;Keywords: actin binding; ATP; coiled coil; hydrolase; nucleotide binding; P-loop; pho
F;1-89/Domain: globular head <HED>
F;89-747/Domain: myosin motor domain homology <MOT>
F;179-186/Region: nucleotide-binding motif A (P-loop)
F;819-2116/Domain: alpha-helical rod <ROD>

Query Match 4.3%; Score 134; DB 2; Length 2116;

Best Local Similarity 20.6%; Pred. No. 3.9; Mismatches 82; Indels 182; Gaps 27;
Matches 123; Conservative 82

QY 16 ANPLQHQSEQRRLPK-IVETESTSMDITIQSQKQFLLK-----IDELAAFSVAV 69
Db 817 ARPLLKRRNFEKEKEIREILELKSNTLDTSTQKLEKLSKDTESNVLQQLQKAEK 876
QY 70 EYFKQFPD-----DLQGHIESIENAIKSLKESNGVLAARN----- 105
Db 877 ETLKAMYDSKDALEAKQKLEIRVEDMESELDK-----KLLENLQNKRSVEEKVRD 930
QY 106 -----NNFHPMLSPRRNNVSVETTVT-VSQPSQEIYVETSNRPEG 145
Db 931 LBEELQEQKLNTLEKLLKKYEESELEEMKRVNDQSQSTISRLEKIKDELQKEVEELTES 990
QY 146 GRMCELMCSKGL-----RKIYANISD-----QAKLMEI----- 175
Db 991 --FSEESKDKGVLEKTRVRLQSELDLTVRLDSEFTKDSKESPMSSARQVSLILLESFLL-----MPD 229
QY 176 -PSALKAKPEAKFVLDICIGFYKQGRRAFTKESPMSSARQVSLILLESFLL-----MPD 229
Db 1049 AETRAKLAQSAANKLQ--GE-YTELNEKFNSEVTARSNVEKSKTLESQVAVNNELDE 1105

QY 230 RKGKVKIESWIKDEAETAAVAMRK-RLMTGG-----LAAAEKWDARGLLLLVACFG 281
DB 1106 EKORDALEK-KKALDAMLEENKQDLESTGEGKSLYDLVKQBSMDMEALANQI-----1159
QY 282 VPSNFRST-DLLDLIRWGSNEIAGALKRSQFLVPMVSGIVGESSIKRGWHIALEAMVYTF 340
DB 1150 --SELQSTIAKLEKIKSTLEGEVA-----RLQGELEBAQLAKSNVEKQKKVEL 1206
QY 341 GMEKFSALVLTSLFKMSKESFERAKRAQSPILA-----FKEAATK-----382
DB 1207 DLEDK-SAQLAET---AAQALDKLKKLEQELSEVOTQLSEANNKNVNSDSTNKHLET 1262
QY 383 -----QAVLSVVMQCMETHKLDPAKELPCWQIKEOI-----VSIKKT 421
DB 1263 SFNNKLEBAEQAKQALEKKRIGLSELEK-HVNEQLEEEKKQKESNKRKVDLEKEV 1320
QY 422 LQDKEMEKEARSLSLMEEAALAKRMYNQOIKRPLSPMPMPVPTSSSYPIYDRS 478
DB 1321 SELKQDIEEVASKAVTEAKNKKSELEIKR-----QVADVSS-----RDKS 1365
RESULT 10
T13564
Microtubule-associated protein homolog - fruit fly (Drosophila melanogaster)
N:Alternate names: hypothetical protein EG:49E4.1
C:Species: Drosophila melanogaster
C>Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
A:Accession: T13564
R:Spanos, L.; Papagiannakis, G.; Siden-Kiamos, I.; Louis, C.
A:Description: Sequencing the distal X chromosome of Drosophila melanogaster.
A:Reference number: Z17689
A:Accession: T13564
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-5327 <SPA>
A:Cross-references: EMBL:AL031128; PIN:CAA20006.1
C:Genetics:
A:Introns: 24/2; 52/3; 104/3; 179/1; 232/1; 1669/3; 2566/1; 4798/3; 5272/1
A>Note: EG:49E4.1
C:Superfamily: Drosophila 576K microtubule-associated protein homolog
Query Match 4.3%; Score 132; DB 2; Length 5327;
Best Local Similarity 20.7%; Pred. No. 20;
Matches 119; Conservative 103; Mismatches 219; Indels 134; Gaps 27;
QY 7 TVAAOPTTA-NPLLQRHS-----EQRRRLPKIVETESTMDITIGQSQPOFLKSID 60
DB 3976 TVTKQSETTLFETLTSKVESKVEVLESSVKQVEKVTQSKQATTTVDSLEQLTKSSE 4035
QY 61 ELA-----AFSAVETFKRQFD-----LQKHESI-----ENADS 92
DB 4036 QLTEIKSVLDTNINVTNLFSTAVETIEKKVQDVTEKVIKATEHVSEHVTTTGESSTET 4095
QY 93 KLESNGVLA--ARNNHQPMSPRNNSVETVTVSQSPQIIVPETSNNKPEGGRV-- 148
DB 4096 SOEKSLLDGTSELRETHITTVGSP-----EFTVICERDEPVLADIKEEDHEHRFSP 4149
QY 149 -----CELMCSKGLRKTIYANISQAKLMBEIIPSALKLAKEPAKFLVDCIGKF-----196
DB 4150 PSDVDKAAIIPQMPRP-LSPREEVAKIVADVAKVLKSKD-----ITDIIPDFDERQLE 4204
QY 197 -VLOGRAFTKESPMSSARQVSLLT-----LESFLAMPDRGKGVKIESWIKDEAETAAV 250
DB 4205 EKLKSTADEESKSTDEKSLSEISVKVEIESKSPDQSGPISIEE--KQIKES-- 4260
QY 251 AWRKRLMTGEGGLAAAEKMDARGLLLVACFGVPSNFRSTDLDDLRMSGNEIAGALKRS 310
DB 4261 --EKALRQGLASSRPESVASQPESVPS---PSQSAASHEHKEVELSESHK-AEKSRRP 4314
QY 311 QFLVPMV-----SGVIESIKG-----MH-TEALEMVTTFGMEKDFSAAL 350

DB 4315 ESVASQVSEKDMKTSRPAASSTQSFSTKEGDETTBSLLHSLTITTTETVETKQMEKSSPES 4374
QY 351 VLTSFLFKMSKESFERAKRAQSPILAFKAATQQLAVLSSVMQCMETHKLDPAKEJPGWQI 410
DB 4375 VTSVTKSTVLS-----SQSTVOLREESTSE--SLSSSLKVEDSSRRSLSSL-----L 4421
QY 411 KEQ-----IVSLKDTQLQDKEME-----EKARSLSLMEE--AALAKRMYNQOIKRPL 457
DB 4422 AEKGGIATNTSLKEDTSASASQLEELLVQSECCSSSEIVSEICTSIAQKS-NKEIKDARE 4480
QY 458 SPMEPPVPTSSSYPIYDRSPSPQRDDDDQDEISA 492
DB 4481 TKVTSQFTTITS-----SATKDDSLKETVA 4505
RESULT 11
S49461
synaptonemal complex protein 1 - mouse
C:Species: Mus musculus (house mouse)
C>Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 05-Nov-1999
A:Accession: S49461; S59599
R:Julien, S.; Luc, M.; Francois, C.
A:Submitted to the EMBL Data Library, October 1994
A:Description: Cloning and sequencing of the murine SCPI cDNA.
A:Reference number: S49461
A:Accession: S49461
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-993 <JUL>
A:Cross-references: EMBL:Z38118; NID:G1360015; PIN:CAA86262.1; PID:G558603
R:Sage, J.; Martin, L.; Curin, F.; Rassoulzadegan, M.
Biochim. Biophys. Acta 1263, 258-260, 1995
A:Title: cDNA sequence of the murine synaptonemal complex protein 1 (SCPI).
A:Reference number: S59599; MUID:96004899; PMID:7548215
A:Accession: S59599
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-993 <SAG>
A:Cross-references: EMBL:Z38118; NID:G1360015; PIN:CAA86262.1; PID:G558603
Query Match 4.2%; Score 131; DB 2; Length 993;
Best Local Similarity 19.8%; Pred. No. 2;
Matches 122; Conservative 97; Mismatches 222; Indels 176; Gaps 29;
QY 13 TTTANPLQRHQSQRRLPKIVETESTMDITIGQSQPOFLKSIDELAAP-----SVA 68
DB 381 TCTLEELL---RTEQQRLE-----KNEDQLKLITVELQKKS---NELEMTKFNKKEVE 429
QY 69 VETPK-----ROPDDLQKHIESIENADSKLESNGVLAARNNNHQPWLSPRNNSV 122
DB 430 LELKNILAEQKLLDEKKQVEKLAELQEKQELTFLLETRKEVH-----DLQE 480
QY 123 ETTVT-----VSQSQEIVPETSNNKPEGGRVCELMCSKGLRKTIYANISQAKLM 172
DB 481 QVTVTKTSEQHYLKQVEEMKTELEKELKNTLTASCDML-----LLENKKEV 528
QY 173 EELPSALKLAKEPAKFLVDCIGKFYQGRFAFTKESPMSSARQVSLTILLESFILMPDRCK 232
DB 529 QASDMALELKHQEDLIINC-----KKQERLLKQIENL-----562
QY 233 GKVKIESWIKDEAETAAVAMRKLMTEGG-----LAAAEKMDARGL-----LL 275
DB 563 --EKEWHLRDELSV---RKEFIQGDVCKLDKSEB-NARSTCEVLEKKEKQMKIL 615
QY 276 LVACFGVPSNF-RSTDLDDI-----IRMSGNEIAGALKRSQFLVPMVSGIVSESIKR 327
DB 616 ESKNNKQVENSCKRIEHLHOENKTKKSSAEIK-QUNAVEIKVSKLELESTKGR 674
QY 328 GMHIEALEMVTTFGMEK-----PSAALVLTSLFKMSKESFERAKRAQSPILAF 376
DB 675 ---FEEMNNVQKIEENKTSSEKGLGEVEKAKATVDEAVKLOKEIDLRCQHKIAEMVAL 731
QY 377 KEAATKQ-----LAVLSVMQCMETHKLDPAKELPGWQIKQIQLVLEKDTQLQD 425

```
Db 732 MEKHKHQYDKIVEERDSELGLYKREOQSAAKTALETLS--NIRNELVSLKK--QLE 786
Qy 426 KEMEEKARSLSLMEAAALAKGMYNQIKRPLSPNEMPPVTSSSVSPYDRGPPSOR-- 483
Db 787 IEKEKEKLMKAKENTAIK---DKDKKIOASLESPEATSWKFD---SKTPSQNIS 839
Qy 484 -----DDQDEISALVSSYLGSTPSPHRSRSPYIMVPLPHGGLGRSVAYEH 532
Db 840 RLSSMSDGSXDNRLNRASAKSIL-PTTV-----TKYTVKTP---TKKSIYQRE- 887

Qy 533 LAPNSYSPGHGRLHRQ 549
Db 888 ---NKYIPTGGSNKKRK 901

RESULT 12
SCPI protein - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 08-Oct-1999
C:Accession: S28061
R:Neuwissen, R.L.J.; Offenberg, H.H.; Dietrich, A.J.J.; Rieswijk, A.; van Iersel, M.; H.
EMBO J. 11, 5091-5100, 1992
A:Title: A coiled-coil related protein specific for synapsed regions of meiotic prophase
A:Reference number: S28061; MUID:93099884; PMID:1464329
A:Accession: S28061
A:Molecule type: mRNA
A:Residues: 1-946 <NEU>
A:Cross-references: EMBL:X67805; NID:G57212; PIDN:CAA48006.1; PID:G57213
C:Genetics:
A:Gene: SCPI
C:Keywords: DNA binding

Query Match 4.2%; Score 129; DB 2; Length 946;
Best Local Similarity 19.1%; Pred. No. 2.5; Mismatches 235; Indels 146; Gaps 25;
Matches 115; Conservative 106;

Qy 13 TTTANPLQHQHSEQRRLPKIVETESTSMDITIGQSKQPFQKSIDELAAF----SVA 68
Db 334 TCTLELL--RTEQRL-----NNEOLKLTWELQKS---SELEMTKFNKNEVE 382

Qy 69 VETFK-----QFDDLOKHIESIENADSKLESNGVLAARNNHPOMLSPNNVSV 122
Db 383 LEBLKTLAEDQKLLDEKKQVEKLAELQGEQELTFLLOTREKH-----DL 431

Qy 123 EITVTVSQSQE-----IVPETSNGKPEGGM--CELMCSKGLRKYYIYANISDAKLMEEIP 176
Db 432 EVQVTVTKTSEHYLKQVEEMTELEKEKLNIELTANSDMI-----LLENKLVQES 485

Qy 177 SALKAKEPAKFVLDICIGFYLGRRAPTKESPMSARQVSLILLESFLMPDRGKGVK 236
Db 486 DMVLELKHQEDIIINCKQBEERMLQIETLEBKEMNLRDELESVRKETFIOGDEVKCKL- 544

Qy 237 IESWIWDEAETAAVARKLMTTEGGLAAEKVDAGLLLVACFGVPSNF--RSTDLIDL 294
Db 545 -----DKSEENARSIEYEV-----KKEQMKILEKNCKMLKQIENKKNIEEL 589

Qy 295 -----IRMSGSN-----ETAGALKRSQFLVPMVSGIVESSIKRGHMI 331
Db 590 HQENKALKKXSAENKQNAVEIKYNKLELELASTQKFEEMI-----NNYQKBEI 641

Qy 332 EALEMVTGMDKFAALVLSFLKMSKESERAKRAQSPFAKAEATVQ----- 383
Db 642 KKTSEELKLGVEK--AKATVDEAVKLEQEDILRCQHKIAENVALMEKHQYDKIVEER 699

Qy 384 ---LAVLSSVMQCMETHKLDPAKELPGWQIQEIVSLEKDTLQLDKEMEERKARSLIMEE 440
Db 700 DSELGCHKREOQSAAKTALETLS--NIRNELVSLKK--QLEVEKEEKLKMEQEN 754

Qy 441 AALAKRMYNQIKRPLSPNEMPPVTSSSVSPYDRGPPSOR-----DDQ 487
Db 755 TAI---LTDKKDKKIOASLESPEATSWKFD---SKTPSQNISRLSSMSDGSKDKNR 807
```

```
Qy 488 DEISALVSSYLGSTPSPHRSRSPYIMVPLPHGGLGRSVAYEHAPNSYSPGHGRLH 547
Db 808 DSLRASAKSILSTTV-----TKYTVKTP---TKKSIYQRE-----NKYLTGGSNKK 852

Qy 548 RQ 549
Db 853 RK 854

RESULT 13
B48666
cell proliferation antigen Ki-67, short form - human
C:Species: Homo sapiens (man)
C:Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 07-May-1999
C:Accession: B48666
R:Schlueter, C.; Duchrow, M.; Wohlenberg, C.; Becker, M.H.G.; Key, G.; Flad, H.D.; Gerdt
J. Cell Biol. 123, 513-522, 1993
A:Title: The cell proliferation-associated antigen of antibody Ki-67: a very large, ubi-
ins.
A:Reference number: A48666; MUID:94043435; PMID:8227122
A:Accession: B48666
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-2897 <SCH>
A:Cross-references: EMBL:X65551
C:Superfamily: kinase interaction domain homology
C:Keywords: alternative splicing; cell cycle control; nucleus; tandem repeat
F:29-91/Domain: Kinase interaction domain homology <KIH>

Query Match 4.1%; Score 127.5; DB 2; Length 2897;
Best Local Similarity 20.0%; Pred. No. 16;
Matches 137; Conservative 92; Mismatches 228; Indels 229; Gaps 32;

Qy 11 OPTTANPLQHQHSEQR--RELPKIVETESTSMDI-----TIGQSKQPOFLK 57
Db 505 ESKTVSTVNSGRTSTFRNQKLPVSKSEETNTEIVEICILKRGQATILQORREGEMK 564

Qy 58 SDELAFAFSAVETFKQFDDLOKHIESIENADSKLESNGVLAARNNHPQ---PMLS 114
Db 565 EIE-----RPFETYKENIELKEN--DEKMA-----MKRSRTWGQKCAPM-- 602

Qy 115 PRNNVSVETVTVSQSQEIVPETS-----KPEGGRMCELMCSKGLRKY 160
Db 603 -----SGLTDLKSLPDTLMKDTARGONLLQTDHAKAPKSEKKITKXPOQSLQPEP 655

Qy 161 IVANISDAKLMEEIPSAKLAKPAKFPVLDICIGF-YLQRRAPTKESPMSARQVSL 219
Db 656 INTPHTKQOL-----KASLGKGVKVELLAVGKFTTSGETTHTHREPAGDKSI-- 706

Qy 220 ILESFLMPD-----RGKGVKIESWIKDEATAA-VAVRKELMTTEGGLAAEKMDAR 271
Db 707 --RTPKESPKOILDPAARVTGKWPRTPKKEAOSLEDLAGFKELFQTPGPSEESMTDEK 764

Qy 272 GLLLVACFGVP-----SNFRSTDL-----LDLIRMSGSNIEIAGALKESQF 312
Db 765 ---TTKIACKSPPEVSDTPTSTKQWPKESLEKADVEEFLALRLKLTPS---AG---KAM 815

Qy 313 LVPVMSGIVESSIK--RGWHIEALEMVTTFQWEDKFAALVLSFLKMSKESERAKKA 370
Db 816 LTPKPAAGGDEKDIKAFMGTPVQKLDLAGT-----LFGSKRLQTPPKKA 859

Qy 371 QS--PLA-FKE-----AATKQ----- 383
Db 860 QALEDLAGFKELFQTPGHTHELVAGKTKTKPCDPSQSDPVDTPSTTKQPKRSIRKADV 919

Qy 384 ----LA-----VLSSVMQCMET-----HKLDPAKELPGWQIQEIVSL 417
Db 920 EGELLACRLNLPISAGKAMHTPKPSVGBEKDIIIFVGTVPVQKLDLTENLTGSKRPQTPKE 979

Qy 418 EKDVTQLDKEMEERKARSLIMEEALAKGMYNQIKRPLSPNEMPPVTSSSVSPYDR 477
Db 980 EAALEDLTGFKELFQTPGHTHEEVAAG-----KTKMPCSSPPESADTPTSTRQP 1032
```

QY 478 SPSQDDDDQDEISAL--VSSYLGPSTSPFRRSRSPFYMVPLPHGGLGRSVAYEHLAP 535
Db 1033 KTFLEKRDVQKELSAKLTQTSGETT---HTDK-----VP---GGEDKSINAFRETAK 1080
QY 536 NSYSPGHGHLRLHROYSPLVHGQRHP 561
Db 1081 QKLDP-----AASVTGSKRHP 1096
RESULT 14
A48666
cell proliferation antigen Ki-67, long form - human
C:Species: Homo sapiens (man)
C:Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 08-Oct-1999
C:Accession: A48666
R:Schueter, C.; Duchrow, M.; Wohlenberg, C.; Becker, M.H.G.; Key, G.; Flad, H.D.; Gerde
J. Cell Biol. 123, 513-522, 1993
A:Title: The cell proliferation-associated antigen of antibody Ki-67: a very large, ubiqui-
tous protein
A:Reference number: A48666; MUID:94043435; PMID:8227122
A:Accession: A48666
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-3256 <SCH>
A:Cross-references: EMBL:X65550; NID:G415818; PIDN:CAA46519.1; PID:G415819
C:Superfamily: Kinase interaction domain homology
C:Keywords: alternative splicing; cell cycle control; nucleus; tandem repeat
F;29-91/Domain: Kinase interaction domain homology <KIH>
Query Match 4.1%; Score 127.5; DB 2; Length 3256;
Best Local Similarity 20.0%; Pred. No. 19; Indels 229; Gaps 32;
Matches 137; Conservative 92; Mismatches 228; Indels 229; Gaps 32;
QY 11 QPTTANPLLRHQSGQR--RELKPIVETESTSMDI-----TIQSKQPOFLX 57
Db 864 EFSKTVSTVNRSGRSTEFNIQKLPESKSEBTNTEIVELKRGKATLLQORREGEMK 923
QY 58 SDELAAPSVAVETFKQDDLOKHIESIENAIKSLKESNGVLAARNNHQ---PMLS 114
Db 924 EIB-----RPPETKYENIELKEN--DEMKKA-----MKSRVTGWQKCAPM-- 961
QY 115 PPRNNVSVETTVTSOPSOEIVPETSND-----KPEGGRMELMCSKGRKY 160
Db 962 -----SDLTDLKSLPDELAKDITARGQNLQATQDHAKAPKSEKIKTKMPCQSLQPEP 1014
QY 161 IVANISQAKLMEEIPSAKLAKPAKFVLDICIGF-VLQGRRAFTKESPMSSARQVSL 219
Db 1015 INTFTTKQOL-----KASLGKVGKVELLAGVFRTISGETTHREPAAGDKSI--- 1065
QY 220 ILESFLLMPD-----RGKGKVKIESMTKDAETAA-VAVKRLMTGEGGLAAAEKMDAR 271
Db 1066 --RTFKESPKQILDPAARVTGMKKWFTPKESBAQSLDLAGFKELFQTPGSEESMTDEK 1123
QY 272 GLLLLVACFGVP-----SNFRSTDLDLIRMGSGNEIAGALKRSQF 312
Db 1124 --TTKIACKSPPPESVDTPTSTKQWPKRSLRADVEEBFLALRLKLTPS---AG---KAM 1174
QY 313 LVPWVSGIVESSIK--RGWHIEALEMVTYFGMEDKFSALVLTSLFKMSKESFERAKRKA 370
Db 1175 LTPKAGGDEKDIKAPMGTPVKLDLAGT-----LPGSKQLQTPKEKA 1218
QY 371 QS--PLA-FKE-----AATKQ----- 383
Db 1219 QALEDLAGFKELFQTPGHTTELVAAGKTKKIPCDSPQDPDVTPTSTKQPKRSIRKADV 1278
QY 384 ----LA---VLSSVMCMET-----HKLDPAKELPGWQKEQIVSL 417
Db 1279 EGGLEACRLMPSAGAMHTPPSPVGEEDKIIIFVGTVPQKLDLTENLTGSKRRPQTPKE 1338
QY 418 EKDTQLDKMEMEEKARSLMEBAALAKRMVNNQIKRPRLSPEMPPVTSSSYPIYDR 477
Db 1339 EAQALDGLTGFELFQTPGHTTEAVAAG-----KTKMPCSSPPESADTPTSTRQP 1391

QY 478 SPSQDDDDQDEISAL--VSSYLGPSTSPFRRSRSPFYMVPLPHGGLGRSVAYEHLAP 535
Db 1392 KTFLEKRDVQKELSAKLTQTSGETT---HTDK-----VP---GGEDKSINAFRETAK 1439
QY 536 NSYSPGHGHLRLHROYSPLVHGQRHP 561
Db 1440 QKLDP-----AASVTGSKRHP 1455
RESULT 15
T19898
aryl hydrocarbon receptor AHR-1 homolog - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jul-2000
C:Accession: T19898; T42398
R:Steward, C.
submitted to the EMBL Data Library, October 1996
A:Reference number: Z19192
A:Accession: T19898
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-602 <WIL>
A:Cross-references: EMBL:Z81048; PIDN:CAB51463.1; GSPDB:GN00019; CESP:C41G7.5
A:Experimental source: Clons C41G7
R:Fowell-Cotman, J.A.; Bradfield, C.A.; Wood, W.B.
Proc. Natl. Acad. Sci. U.S.A. 95, 2844-2849, 1998
A:Title: Caenorhabditis elegans orthologs of the aryl hydrocarbon receptor and its heter-
odimeric ligand
A:Reference number: Z22156; MUID:98169459; PMID:9501178
A:Accession: T42398
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-602 <POM>
A:Cross-references: EMBL:AF039570; NID:G2828113; PIDN:AAC00000.1; PID:G2828114
C:Genetics:
A:Gene: ahr-1; C41G7.5
A:Map position: 1
A:Introns: 13/2; 33/2; 76/1; 210/1; 234/3; 270/2; 303/2; 338/1; 386/2; 436/2; 541/1; 557
Query Match 4.1%; Score 126.5; DB 2; Length 602;
Best Local Similarity 19.0%; Pred. No. 1.9; Indels 171; Gaps 29;
Matches 120; Conservative 99; Mismatches 243; Indels 171; Gaps 29;
QY 5 PPTVAAQPTTANPLLRHQSGQRRELKPIVETESTSMDITTIQSKQPOFLKSIDELAA 64
Db 17 PP-----KQLTWNTN-----SKRHERLING-ELETVAMLLPYDSST----ISRLDKLSV 60
QY 65 FSVAVETFKQDDLOKHIESIENAIKSLKESNGVLAARNNHQ---HOPMLSPFRNVSV 122
Db 61 LRLAY-----SFLQCKAHFOACLH--NSQFLSAGFPMSTHSYSYQHPPTI--PFSNKVPT 111
QY 123 ETTVTVSOPSOEIVPETSND-----KPEGGRMCEL-----MCSKGLRKYI----- 161
Db 112 IFDLRIGTQPMLD--PESNPFELSLKSLGGFVLVNDNGEIIYVASENVENYLGHQSDVL 169
QY 162 -----YANISQAKLMEEIPSAKLAKPAKFVLDICIGKFIQGRRAFTKESPMASARQ 215
Db 170 HQPVYDLIHSERDDIRQQLDSNFHPTTSSASNQD-----VFAPQNSKYLERN 218
QY 216 VSLILLESFLAMPGRGKGVKIESWIKDEAETAFAVAVKRLMTGEGGLAAAEK--DARGL 273
Db 219 VN-----AFRCILDTNCGFLID-----NRGLKLSLHGLPSSVYVWGRATSGP 261
QY 274 LLLVACFGVP-----SNFRSTDLDLIRMGSGNEIAGALKRSQFLVPMVSGIVESSIKR 327
Db 262 VLGMICVCTPFPVPESTSLASEDMI----LTKHQLDGALVSMQDKVYEMLEIDETDLP 317
QY 328 GMHIEALEMVTYFGMEDKFSALVLTSLFKMSKESFERAKRKAQSLAPAKFAATKQLAVL 387
Db 318 PLV-----NLVH---VEDA-----VCVAEAKKAIKNGSGLLVYRLVTKTTRTY 360
QY 388 SSVVMCMETHKLDPAKELPGWQKEQIVSLEKDTQLDKMEMEEKARSLMEBAALAKRM 447

Db 361 FVSSCEMFYKSKPESI---GLTHRLLEVEGTWLEKRTLKAKLLSFDDSF----- 411
Qy 448 YNQQIKRPLSPMEMPPVTSSYSPIYRDRSPSQDDDDDEISALVSSYLGSTSP-- 505
Db 412 ---LQSPR-----NLQSTAALPL-----PSVLKDDQDCLEPSTNSLFPSPVPTP 454
Qy 506 -----HRSRRSPYMWPLPHGGLGRSVYAYEHLAPNSYSPGHGRLHRCYSPSLVHGQ 558
Db 455 TTTKANRRKENSHEIVTIP-----SIPITHEFDMQMFDPSPWNHGVH---PPAWPHDV 505
Qy 559 RHPLOYSPPIHGQQQLPYGIGIQRVYRHSPEERY 591
Db 506 YHLTOYPP-----TYPHPGTVGY 524

Search completed: August 14, 2003, 10:40:50
Job time : 46 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 14, 2003, 10:28:09 ; Search time 25 seconds
(without alignments)
1145.571 Million cell updates/sec

Title: US-09-890-475-1

Perfect score: 3104

Sequence: 1 MSNYPPTVAQPTTTANPLL.....RYLGLSNQSPRSNSLDPK 609

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3104	100.0	609	1	FRI_ARATH
2	144	4.6	724	1	Q9fdw0 arabidopsis
3	136	4.4	1722	1	P29375 mus musculus
4	134	4.3	2116	1	P08799 dictyosteli
5	131	4.2	993	1	SCPI_MOUSE
6	129	4.2	987	1	SCPI_MOUSE
7	127.5	4.1	3256	1	K167_HUMAN
8	126.5	4.1	1920	1	PCT2_MOUSE
9	126	4.1	694	1	SR72_CAEEL
10	126	4.1	2230	1	GOG4_HUMAN
11	125.5	4.0	6895	1	SNE3_HUMAN
12	125	4.0	1128	1	BEM3_YEAST
13	124.5	4.0	1197	1	CLNG_HUMAN
14	123.5	4.0	5120	1	PCLO_CHICK
15	123	4.0	485	1	BC13_HUMAN
16	122.5	3.9	1336	1	W146_HUMAN
17	122.5	3.9	2212	1	T230_HUMAN
18	122.5	3.9	962	1	IF3A_MAIZE
19	122	3.9	1325	1	G160_MOUSE
20	120.5	3.9	1875	1	MLP_YEAST
21	118.5	3.8	2044	1	SIF2_DROME
22	118.5	3.8	2064	1	SIF1_DROME
23	117.5	3.8	1060	1	EG51_XENLA
24	117.5	3.8	2842	1	APC_RAT
25	116	3.7	1087	1	E4L3_HUMAN
26	115	3.7	1499	1	RHG5_HUMAN
27	115	3.7	1690	1	C190_DROME
28	114.5	3.7	782	1	YQ91_CAEEL
29	114.5	3.7	943	1	YLW5_CAEEL
30	114.5	3.7	1940	1	MYH3_RAT
31	114.5	3.7	1972	1	MYH6_HUMAN
32	114.5	3.7	2404	1	SON_MOUSE
33	114.5	3.7	4473	1	PLEI_CRIGR

34	114	3.7	1125	1	YB62_SCHPO
35	114	3.7	1433	1	Y310_HUMAN
36	114	3.7	1972	1	MYHB_MOUSE
37	113.5	3.7	1025	1	ADA1_YEAST
38	113.5	3.7	1505	1	CUT2_HUMAN
39	113.5	3.7	1861	1	MAP2_RAT
40	113.5	3.7	1940	1	MYH3_CHICK
41	113.5	3.7	1940	1	MYH3_HUMAN
42	113.5	3.7	2349	1	TPR_HUMAN
43	113	3.6	554	1	RCN1_RICPR
44	113	3.6	1790	1	USO1_YEAST
45	113	3.6	1938	1	MYH4_RABIT

ALIGNMENTS

RESULT 1

FRI_ARATH
ID FRI_ARATH STANDARD; PRT; 609 AA.
AC Q9FDW0; O65274;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE FRIGIDA protein.
GN FRI OR Ar4G0650 OR F6N23.25.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosid II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A., AND VARIANTS EDI-0; SF-2 AND SHAKHDARA.
RC STRAIN=cv. H51;
RX MEDLINE=20485641; PubMed=11030654;
RA Johanson U., West J., Lister C., Michaels S., Amasino R.M., Dean C.;
RT "Molecular analysis of FRIGIDA, a major determinant of natural
variation in Arabidopsis flowering time.";
RL Science 290:344-347(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=20083488; PubMed=10617198;
RA Mayer K.F.X., Schueller C., Wambutt R., Murphy G., Volckaert G.,
Pohl T., Duesterhoeft A., Stiekema W., Entian K.-D., Terryn N.,
Harris B., Anseorge W., Brandt P., Grivell L., Rieger M.,
Weichselgartner M., de Simone V., Obermaier B., Mache R., Mueller M.,
Kreiss M., Delsen M., Puigdomenech P., Watson M., Schmidheini T.,
Reichert B., Portetelle D., Perez-Alonso M., Boutry M., Bancroft I.,
Vos P., Hoheisel J., Zimmermann W., Wedler H., Ridley P.,
Laingham S.-A., McCullagh B., Bilham L., Robben J.,
Van der Schueren J., Grymonprez B., Chuang Y.-J., Vandenbussche F.,
Braeken M., Weltjens I., Voet M., Bastiaens I., Aert R., Defoor E.,
Weitzengerter T., Bothe G., Ransperger U., Hilbert H., Braun M.,
Holzer E., Brandt A., Peters S., van Staveren M., Dirkse W.,
Moolman P., Klein Lankhorst R., Rose M., Hauf J., Koetter P.,
Bernseiser S., Hempel S., Feldpausch M., Lamberth S., Van den Daele H.,
De Keyser A., Buyschaert C., Gielen J., Villarroel R., De Clercq R.,
Van Montagu M., Rogers J., Cronin A., Quail A., Bray-Allen S.,
Clark L., Doggett J., Hall S., Kay M., Lennard N., McIlroy K., Mayes R.,
Pettett A., Rajandream M.A., Lyne M., Benes V., Reimann S.,
Borkova D., Bloeker H., Schaefer M., Grimm M., Loehner T.-H.,
Dose S., de Haan M., Maare A., Schaefer M., Mueller-Auer S.,
Gabel C., Fuchs F., Fartmann B., Granderath K., Dauner D., Herzl A.,
Neumann S., Atgiriou A., Vitale D., Liguori R., Piravandi E.,
Massenet O., Quigley F., Clabaud G., Muendlein A., Felber R.,
Schabl S., Hiller R., Schmidt W., Lechary A., Aubourg S.,
Chefor F., Cooke R., Berger C., Monfort A., Casacuberta E.,
Gibbons T., Weber N., Vandenbol M., Barges M., Terol J., Torres A.,
Perez-Perez A., Purnelle B., Bent E., Johnson S., Tacon D., Jesse T.,
Heijnen L., Schwarz S., Scholler P., Heber S., Francis P., Bielke C.,
Frishman D., Haase D., Lemcke K., Mewes H.-W., Stocker S.,
Zaccaria P., Bevan M., Wilson R.K., de la Bastide M., Habermann K.,

RA Parnell L., Dedhia N., Gnoj L., Schutz K., Huang E., Spiegel L.,
RA Sekhon M., Murray J., Sheet P., Cordes M., Abu-Threideh J.,
RA Stoneking T., Kalicki J., Graves T., Harmon G., Edwards J.,
RA Latreille P., Courtney L., Cloud J., Abbott A., Scott K., Johnson D.,
RA Minx P., Bentley D., Fulton B., Miller N., Greco T., Kemp K.,
RA Kramer J., Fulton L., Mardis E., Dante M., Pepin K., Hillier L.,
RA Nelson J., Spieth J., Ryan E., Andrews S., Geisel C., Layman D.,
RA Du H., Ali J., Berghoff A., Jones K., Drone K., Cotton M., Joseph C.,
RA Antoniou B., Zidanic M., Strong C., Sun H., Lamar B., Yordan C.,
RA Ma P., Zhong J., Preston R., Vil D., Shekher M., Matero A., Shah R.,
RA Swaby J.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Till S.,
RA Granat S., Shohdy N., Hasegawa A., Hameed A., Lodhi M., Johnson A.,
RA Chen E., Maria M., Martienssen R., McCombie W.R.,
RT "sequence and analysis of chromosome 4 of the plant Arabidopsis
thaliana";
RL Nature 402:769-777(1999).
CC -!- FUNCTION: Required for the regulation of flowering time in the
late-flowering phenotype. Increases RNA levels of 'FLOWERING LOCUS
C'. Variants with an early-flowering phenotype (including cv.
Columbia, cv. Landsberg Erecta and cv. Wassilewskija) show loss-
of-function mutations of FRI.
CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -!- CAUTION: The protein sequence is incorrectly predicted in the
genomic sequence of cv. Columbia, but as the gene is defective in
this cultivar it cannot be predicted at protein level correctly.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch)
CC -----
DR EMBL; AF228500; AAC23415.1; -;
DR EMBL; AF228499; AAC23414.1; -;
DR EMBL; AF058919; AAC13626.1; ALT SEQ.
DR EMBL; AL161472; CAB80874.1; ALT SEQ.
KW Flowering; Coiled coil; Developmental protein; Nuclear protein.
FT DOMAIN 409 440 COILED COIL (POTENTIAL).
FT DOMAIN 55 55 COILED COIL (POTENTIAL).
FT VARIANT 79 79 F -> I (IN CV. SHAKHARA).
FT VARIANT 146 146 L -> I (IN CV. SF-2).
FT VARIANT 148 148 G -> E (IN CV. EDI-O).
FT VARIANT 148 148 M -> I (IN CV. EDI-O).
SQ SEQUENCE 609 AA; 68443 MW; FE8572DF3E6726F CRC64;

Query Match 100.0%; Score 3104; DB 1; Length 609;

Best Local Similarity 100.0%; Pred. No. 6.5e-182; Indels 0; Gaps 0;

Matches 609; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSNYPPTAAQPTTTANPLLRHQSEQRRELPKIVETESTSMDITIGSKQPQFLKSID 60
DB 1 MSNYPPTAAQPTTTANPLLRHQSEQRRELPKIVETESTSMDITIGSKQPQFLKSID 60
QY 61 ELAAPSVAVETKRPQDDLOKHIEIENAIKSLKESGVLAARNPHQPLSPRRNV 120
DB 61 ELAAPSVAVETKRPQDDLOKHIEIENAIKSLKESGVLAARNPHQPLSPRRNV 120
QY 121 SVETTVTSQPSQEIIVPETSINPEGRMCELMCSKGLKRYANIISDAKLMEIIPSAK 180
DB 121 SVETTVTSQPSQEIIVPETSINPEGRMCELMCSKGLKRYANIISDAKLMEIIPSAK 180
QY 181 LAKEPAKFLVDCIGFYLOGRAFTKESPMSSARQVSLILLESFLMPDRGKGKIESW 240
DB 181 LAKEPAKFLVDCIGFYLOGRAFTKESPMSSARQVSLILLESFLMPDRGKGKIESW 240
QY 241 IKDEATAAVAKRLMTGGGLAAEKMDARGLLLLVACFGVPSNFRSTDLDLIRMSG 300
DB 241 IKDEATAAVAKRLMTGGGLAAEKMDARGLLLLVACFGVPSNFRSTDLDLIRMSG 300
QY 301 NETAGALKRSQFLVPMVSGIVGESSIKRGHIEALEMWTYFGMEDKFSALVLTFLKMSK 360

DB 301 NETAGALKRSQFLVPMVSGIVGESSIKRGHIEALEMWTYFGMEDKFSALVLTFLKMSK 360
QY 361 ESFRAKRAKQSPFLAFKFAATKQALVLSVMQCMETHKLDPAKELPGWQIKQIVSLEKD 420
DB 361 ESFRAKRAKQSPFLAFKFAATKQALVLSVMQCMETHKLDPAKELPGWQIKQIVSLEKD 420
QY 421 TLQDKEMEKEKARSLSMEEAALAKRMYNQOIKRPLSPMEWPPVTSSSYSPYDRSP 480
DB 421 TLQDKEMEKEKARSLSMEEAALAKRMYNQOIKRPLSPMEWPPVTSSSYSPYDRSP 480
QY 481 SORDDDODEISALVSSYLGSTSPFPHRSRSPYMWPLPHGGGLGRSVYAYEHLAPNSYP 540
DB 481 SORDDDODEISALVSSYLGSTSPFPHRSRSPYMWPLPHGGGLGRSVYAYEHLAPNSYP 540
QY 541 GHGHLRHQRQSPSLVHGQRHPLQSPPIHQOQQLPGVQIRVYHRSSEERYLGLSNQRP 600
DB 541 GHGHLRHQRQSPSLVHGQRHPLQSPPIHQOQQLPGVQIRVYHRSSEERYLGLSNQRP 600
QY 601 RNSNSLDPK 609
DB 601 RNSNSLDPK 609
RESULT 2
ID HMNR HUMAN STANDARD; PRT; 724 AA.
AC O75330; Q92767;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hyaluronan mediated motility receptor (intracellular hyalurononic acid
binding protein) (Receptor for hyaluronan-mediated motility) (CD168
antigen).
GN HMNR OR IHABP OR RHAMM.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., ALTERNATIVE SPLICING, AND CHARACTERIZATION.
RC TISSUE=Breast carcinoma;
RX MEDLINE=98264864; PubMed=9601098;
RA Assmann V., Marshall J.F., Fieber C., Hofmann M., Hart I.R.,
RT "The human hyaluronan receptor RHAMM is expressed as an intracellular
protein in breast cancer cells."
RL J. Cell Sci. 111:1685-1694(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Breast;
RX MEDLINE=97045629; PubMed=8890751;
RA Wang C., Entwistle J., Hou G., Li Q., Turley E.A.;
RT "The characterization of a human RHAMM cDNA: conservation of the
hyaluronan-binding domains."
RL Gene 174:299-306(1996).
CC -!- FUNCTION: INVOLVED IN CELL MOTILITY. WHEN HYALURONAN BINDS TO
HMNR, THE PHOSPHORYLATION OF A NUMBER OF PROTEINS, INCLUDING THE
FOCAL ADHESION KINASE OCCURS. MAY ALSO BE INVOLVED IN CELLULAR
TRANSFORMATION AND METASTASIS FORMATION, AND IN REGULATING
EXTRACELLULAR-REGULATED KINASE (ERK) ACTIVITY.
CC -!- SUBUNIT: SUBUNIT OF THE HARC COMPLEX.
CC -!- SUBCELLULAR LOCATION: CELL SURFACE AND CYTOPLASMIC (BY
SIMILARITY).
CC -!- ALTERNATIVE PRODUCTS:
Event=Alternative splicing; Named isoforms=2;
CC Name=A;
CC IsoId=O75330-1; Sequence=Displayed;
CC Name=B;
CC IsoId=O75330-2; Sequence=VSP_004286;
CC -!- TISSUE SPECIFICITY: EXPRESSED IN BREAST CANCER CELL LINES AND IN
NORMAL BREAST TISSUE.
CC -!- DATABASE: NAME=PROW; NOTE=PROW 2:76-84(2001);
CC WWW="http://www.ncbi.nlm.nih.gov/prow/guide/802868666.g.htm".
CC -----

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC GO:0007048; P:oncogenesis; TAS.
 CC HYALURONIC ACID; Alternative splicing; Repeat; Glycoprotein; Antigen.
 CC FT DOMAIN 635 645
 CC FT CARBOHYD 133 133
 CC FT CARBOHYD 477 477
 CC FT CARBOHYD 567 567
 CC FT CARBOHYD 588 588
 CC FT VARSPLIC 75 90
 CC FT CONFLICT 75 75
 CC FT CONFLICT 103 103
 CC FT CONFLICT 277 277
 CC FT CONFLICT 298 298
 CC FT CONFLICT 322 322
 CC FT CONFLICT 330 332
 CC FT SEQUENCE 724 AA; 84031 MW; EAG8AD6D2A626926 CRC64;
 Query Match 4.6%; Score 144; DB 1; Length 724;
 Best Local Similarity 20.0%; Pred. No. 0.18;
 Matches 112; Conservative 91; Mismatches 220; Indels 136; Gaps 21;
 QY 34 KIVETSTNDITIGSQFQPKSIDELAARSAVETFKQFDLQKHIESIENADS- 92
 Db 215 KLVSIKEKID---EKSETEKLEVEIEISCASDOVEKYKIDIALEENLKEKQDEILSL 271
 QY 93 --KLESGNVLA-----ARNNHQPMSPRNNVSVEITVTVSQ 130
 Db 272 KQLEENIVILSKQVEDLVNKKLEKEKEDHVRNREHNVLENMQL----- 321
 QY 131 PSQEIYVETSNPEGRMCMSCGLKRYIYANISDQAKL---MEETPSALKAKPAK 187
 Db 322 -KQKFILEQOER-EKIQKQELQIDSLQKELSSHLQKLCFOEVMVKEKLFEBELK 379
 QY 188 FVLDCIGKYL---QGRRAFTKSPSSARQVSLILLESFILMPDRGKVKLISWIKOE 244
 Db 380 QTLDLQKQKEEQEQAERLVKOLEEAKSRABELKLLBEKL-----KQK-----EAE 426
 QY 245 AETAAVAKRRLMTEGLAAAEKMDARGILLLVACFGVPSNFRS----- 288
 Db 427 LEKSSAA---HTQATLLQEKYDS---MVQSLQEDVTAQFESYKALTASEIEDLKLEN 477
 QY 289 TDLLDIRMGSN-----EIALGKRSQFLVPMVSGIVESSIKGMHIALEWVYTFGM 342
 Db 478 SSLQKAAKAGKNAEDVQHQILATESSSNOEYRMLDLQTKSALKEIKETISVFLQKI 537
 QY 343 EDKFSALVLTSLFKSKESPERAKRKAQSPAPFAEATKQLAVLSSVMQCMETHKLDP 402
 Db 538 TD-----LQNLQKQEDDFRQKLEDESGRAEKENTAEI-----T 573
 QY 403 KELPGWQ-KEQIVSLQEKD-TLQLDK-EMEEKARSISLMEEAALAKRYN----- 449
 Db 574 EEINKVRLLYEELYNKTKPQLQDAFEVQKQ-----LILNEHGAQEQQLKIRDSYAKLL 629
 QY 450 --QCIKEPRLSPEMPPVTSVSSSPYRDRSPFSORDDDQDSIALVSSYLQ-----PST 502
 Db 630 GHNLKQKIKHVKLKDENSQKSEVSKURCOLAKKKQSETKLQELNKLVLGIKHFDESK 689
 QY 503 SFPHRSRSPYMPVLPFHG 521
 Db 690 AFHESKENFALTKPLKEG 708

RESULT 3
 RB22 HUMAN STANDARD; PRT; 1722 AA.
 ID RB22 HUMAN
 AC P29375;
 DT 01-DEC-1992 (Rel. 24, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Retinoblastoma-binding protein 2 (RBBP-2).
 GN RBBP2 OR RBP2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OC NCBI_TaxID:9606;
 RP [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94020841; PubMed=8414517;
 RA Fattaey A.R., Helin K., Dembski M.S., Dyson N., Harlow E.,
 RA Vuocolo G.A., Hanobik M.G., Haskell K.M., Oliff A., Defeo-Jones D.,
 RA Jones R.E.;
 RA "Characterization of the retinoblastoma binding proteins RBP1 and
 RT RBP2.";
 RL Oncogene 8:3149-3156(1993).
 RP [2]
 RP SEQUENCE OF 1102-1569 FROM N.A.
 RX MEDLINE=91312450; PubMed=1857421;
 RA Defeo-Jones D., Huang P.S., Jones R.E., Haskell K.M., Vuocolo G.A.,
 RA Hanobik M.G., Huber H.E., Oliff A.;
 RT "Cloning of cDNAs for cellular proteins that bind to the
 RT retinoblastoma gene product.";
 RL Nature 352:251-254(1991).
 CC -!- FUNCTION: INTERACTS WITH THE VIRAL PROTEIN-BINDING DOMAIN OF THE
 CC RETINOBLASTOMA PROTEIN.
 CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).
 CC -!- SIMILARITY: TO THE C-TERMINAL DOMAIN OF THE XE169 PROTEIN AND
 CC THE JUMONJI PROTEIN.
 CC -!- SIMILARITY: Contains 3 PHD-type zinc fingers.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC SUBFRAGMENT (82).

CC -1- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING

CC CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,

CC CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.

CC -1- PTM: PHOSPHORYLATION INHIBITS THICK FILAMENT FORMATION AND REDUCES

CC THE ACTIN-ACTIVATED ATPASE ACTIVITY.

CC -1- MISCELLANEOUS: DICTYOSTELIUM MYOSIN II HAS NO K(2)EDTA ATPASE

CC ACTIVITY, PERHAPS CORRELATED WITH THE ABSENCE OF A CYS AT THE SH-1

CC POSITION (688).

CC -1- SIMILARITY: Contains 1 myosin-like globular head domain.

CC -1- SIMILARITY: Contains 1 IQ domain.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

CC EMEL; M14628; AAA33227.1; --

CC FIR; A26655; A26655.

CC PDB; 1MMA; 03-DEC-97.

CC PDB; 1MMD; 17-AUG-96.

CC PDB; 1MMG; 03-DEC-97.

CC PDB; 1MMN; 03-DEC-97.

CC PDB; 1MND; 17-AUG-96.

CC PDB; 1MNE; 17-AUG-96.

CC PDB; 1VOM; 23-DEC-96.

CC PDB; 1LVK; 28-JAN-98.

CC PDB; 1DOX; 20-DEC-00.

CC PDB; 1DOY; 20-DEC-00.

CC PDB; 1DOZ; 20-DEC-00.

CC PDB; 1D1A; 20-DEC-00.

CC PDB; 1D1B; 20-DEC-00.

CC PDB; 1D1C; 20-DEC-00.

CC PDB; 1FMV; 20-DEC-00.

CC PDB; 1FMW; 20-DEC-00.

CC PDB; 1GBX; 17-JAN-01.

CC PDB; 1JMY; 07-NOV-01.

CC PDB; 1JX2; 07-NOV-01.

CC DictyDb; DD01008; mhca.

CC InterPro; IPR000048; IQ_region.

CC InterPro; IPR001609; myosin_head.

CC InterPro; IPR004009; Myosin_N.

CC Pfam; PF00612; IQ; 2.

CC Pfam; PF00633; myosin_head; 1.

CC Pfam; PF02736; Myosin_N; 1.

CC PRINTS; PR00193; MYOSINHEAVY.

CC ProDom; PD000355; myosin_head; 1.

CC SMART; SM00015; IQ; 1.

CC SMART; SM00242; MYSC; 1.

CC PROSITE; PS00096; IQ; 1.

CC Calmodulin-binding; Actin-binding; ATP-binding; 3D-structure;

CC Myosin; Coiled coil; Alkylation; Phosphorylation.

CC -----

CC DOMAIN 1 761

CC DOMAIN 762 791

CC DOMAIN 817 2116

CC NP_BIND 179 186

CC DOMAIN 638 660

CC DOMAIN 738 752

CC MOD_RES 130 130

CC MOD_RES 678 678

CC MOD_RES 1823 1823

CC MOD_RES 1833 1833

CC MOD_RES 2029 2029

CC TURN 3 5

CC TURN 7 8

CC TURN 10 15

CC TURN 25 28

CC TURN 34 37

CC TURN 41 42

CC TURN 48 55

FT STRAND 59 63

FT STRAND 69 73

FT HELIX 74 76

FT STRAND 78 79

FT HELIX 83 85

FT TURN 86 87

FT STRAND 90 90

FT STRAND 91 93

FT HELIX 99 110

FT TURN 111 113

FT STRAND 116 119

FT TURN 120 121

FT STRAND 122 126

FT STRAND 137 142

FT TURN 143 145

FT HELIX 148 150

FT HELIX 155 159

FT TURN 170 171

FT STRAND 173 179

FT TURN 181 182

FT HELIX 185 200

FT HELIX 210 226

FT STRAND 227 228

FT TURN 234 235

FT STRAND 240 247

FT TURN 249 250

FT STRAND 253 261

FT HELIX 265 268

FT TURN 269 289

FT TURN 273 274

FT STRAND 278 278

FT HELIX 279 287

FT HELIX 290 296

FT TURN 297 297

FT HELIX 301 303

FT TURN 305 307

FT TURN 316 317

FT HELIX 320 334

FT TURN 335 335

FT HELIX 338 355

FT TURN 356 356

FT STRAND 360 360

FT STRAND 368 368

FT HELIX 373 382

Query Match 4.3%; Score 134; DB 1; Length 2116;

Best Local Similarity 20.6%; Pred. No. 3.1;

Matches 123; Conservative 82; Mismatches 210; Indels 182; Gaps 27;

QY 16 ANPLQHQHQEQRRRLPK-IVETESTMDITIGSKQPOFLKS-----IDELAAPSVAV 69

DB 817 ARPLKRRNFKEIKEREILELKSNTLDTSTTKDKLEKSKDKTESNVLQOLKAEK 876

QY 70 EYFKRQFD-----DLQKHIESIENAIKSLGNSGVVLAARN----- 105

DB 877 ETLKMYDSKDALEAKRELEIRVEDMESELDK-----KALENLQNKRSVEEKVRD 930

QY 106 -----NNFHQWMLSPRRNVSVETTVT-VSQPQEIIVPETSNTPEG 145

DB 931 LBEELQEQKRLNTLEKLKKYEELEEEMKRVNDQSDTISRLEKIKDELQKEVEELTES 990

QY 146 GRMCELMCSKGL---RKVIYANISD-----QAKLMBEI----- 175

DB 991 --FSESKDKGVLEKTRVRLQSELDDLTVRLDSTKOKSELLROKKLEBELKQVQAL 1048

QY 176 -PSALKLAKPAKFLVDCIGKPYLQGRPAFTKESPMGSAQVSLILLESFIL-----MPD 229

DB 1049 AETAAKLAQEAANKKLQ--GE-YTELNEKFNSEVTARSNVEKSKTLESQLVAVNNELDE 1105

QY 230 RGKGKVKIESWIKDEAETAATAVAWRK-RLWTEGG-----LAAAEKMDARGLLLVACFG 281

DB 1106 EKNRDALEK--KKVALDAMLEEMKQDLESTGGEEKKSLYDLKVKQESDMEALRNQI----- 1159

QY 282 VPSNFRST-DLLDLIRMSGNEIAGALKRSQFLVPMVSGIVSESSIKRMHIEALEMVTFF 340
DB 1160 --SELQSTTIALEKIKSTLEGEVA-----RLQGELEAQLAKSNVYKQKKVEL 1206
QY 341 GMEDEFSAALVLTSLFKSGSEFERAKKAOSPLA-----FKEAATK----- 382
DB 1207 DIEDK-SAGLAET---AAQALDKLKKLQELSEVOTQISEANNKVNDSSTNKHLET 1262
QY 383 -----QLAVLSVVMQCMETHKLDPAKELPGWQIKQI-----VSLKOT 421
DB 1263 SFNNLKLLEAQAQAKALEKKRLGLESELK--HYNEQLSEEEKQKESNEKRVLDLEKEV 1320
QY 422 LOLDKEMEBAKSLISMEBAALAKRMWQIKRPLSPMPPVTSSSYSPYDRDS 478
DB 1321 SELKQIEBEVAKVTEAKNKKESELDEIKR-----QYADVSS-----RDXS 1365

RESULT 5
SCPI_MOUSE STANDARD; PRT; 993 AA.
AC Q62259; O09205; P70192; Q62329;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Synaptonemal complex protein 1 (SCP-1 protein).
GN SYCP1 OR SCP1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CBA; TISSUE=Testis;
RX MEDLINE=96004899; PubMed=7548215;
RA Sage J., Martin L., Cuzin F., Rassoulzadegan M.;
RL "cDNA sequence of the murine synaptonemal complex protein 1 (SCP1).";
RL Biochim. Biophys. Acta 1263:258-260(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Swiss; TISSUE=Testis;
RA Kerr S.M., Taggart M.H., Lee M., Cooke H.J.;
RL Submitted (APR-1995) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 1-149 FROM N.A.
RC STRAIN=C57BL/6;
RA Sage J., Li Y., Martin L., Mattei M.-G., Guenet J.-L., Liu J.G.,
RL Hoog C., Cuzin F., Rassoulzadegan M.;
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE OF 95-787 FROM N.A.
RC STRAIN=ICR; TISSUE=Testis;
RA Teuchida J., Nishina Y., Nozaki M., Uchida K., Nishimune Y.;
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: MAJOR COMPONENT OF THE TRANSVERSE FILAMENTS OF
CC SYNAPTONEMAL COMPLEXES (SCS), FORMED BETWEEN HOMOLOGOUS
CC CHROMOSOMES DURING MEIOTIC PROPHASE.
CC -!- SUBCELLULAR LOCATION: NUCLEAR. IN TRIPARTITE SEGMENTS OF
CC SYNAPTONEMAL COMPLEXES, BETWEEN LATERAL ELEMENTS IN THE NUCLEUS.
CC IS FOUND ONLY WHERE THE CHROMOSOME CORES ARE SYNAPSED. ITS N-TERMINUS
CC IS FOUND TOWARDS THE CENTRE OF THE SYNAPTONEMAL COMPLEX WHILE THE
CC C-TERMINUS EXTENDS WELL INTO THE LATERAL DOMAIN OF THE
CC SYNAPTONEMAL COMPLEX (BY SIMILARITY).
CC -!- DOMAIN: CONSISTS OF AN ALPHA-HELICAL STRETCH OF 700 AA RESIDUES,
CC FLANKED BY N- AND C-TERMINAL GLOBULAR DOMAINS. THE C-TERMINAL
CC DOMAIN HAS DNA-BINDING CAPACITY (BY SIMILARITY).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>)

RESULT 6

CC or send an email to license@isb-sib.ch.
CC -----
DR EMBL; Z38118; CRA86262.1; -
DR EMBL; L41069; AA64514.1; ALT INIT.
DR EMBL; U62864; AAC53335.1; -
DR EMBL; U62860; AAC53335.1; JOINED.
DR EMBL; U62861; AAC53335.1; JOINED.
DR EMBL; U62862; AAC53335.1; JOINED.
DR EMBL; U62863; AAC53335.1; JOINED.
DR EMBL; D88539; BAA13639.1; -
DR PIR; S49461; S49461.
DR MGD; MGI:105931; Sycp1.
KW Nuclear protein; Meiosis; Cell division; Phosphorylation;
KW DNA-binding; Coiled coil.
FT DOMAIN 12 97
FT DOMAIN 104 815
FT DOMAIN 114 117
FT DOMAIN 597 700
FT DOMAIN 898 901
FT DOMAIN 978 986
FT CONFLICT 527 527
SQ SEQUENCE 993 AA; 115962 MW; 1A4FA790D64FAFE6 CRC64;
Query Match 4.2%; Score 131; DB 1; Length 993;
Best Local Similarity 19.8%; Pred. No. 1.7;
Matches 122; Conservative 97; Mismatches 222; Indels 176; Gaps 29;
QY 13 TTTANPLLRHCHSQRRELPLKIVETESTMDITIGOSKOPFLKSIDELAAF----SVA 68
DB 381 TCTLEELL---RTEQORLE-----KNEDQLKLITVELQKS---NELEWTKFNKKEVE 429
QY 69 VETFK-----QFDDLQKHIESIENAIKSLNSGVVLAARNNNHQPMLSPRRNVSV 122
DB 430 LEELKNLILAEQDKLLDEKKQVEKLAELOEKEOELTFLLETRKEVH-----DLQE 480
QY 123 ETTVT-----VSQSQEIVPETSNNKPEGRMCMLCKSLKRYIYANISDOAKLM 172
DB 481 QVTVTISQHYLKQVEEMKTELEKLNTELITASCDML-----LLENKKEV 528
QY 173 BEIPSALKLAKEPAKFLVDCIGFYLOGRRAFTKBSPMSARQVSLILLESFLMLPDRGK 232
DB 529 QEASDMALELKKHQEDIINC-----KKOERLLKQIENL-----562
QY 233 GKVIESWIKDEAETAAVANVKRLMTEGG-----LAAAEKMDARGL-----LL 275
DB 563 --EEKEMHRLDELSV---RKEFIQGGDEVCKLDKSEE--NARSIECEVLKKEKQMKIL 615
QY 276 LVACFGVPSNF--RSTDLIDL-----IRMSGNEIAGALKRSQFLVPMVSGIVSESSIKR 327
DB 616 ESKCNLKKQVENKSKNIEELHQBENKTLKKKSAEIK-QLNAYEIKVSKLELELESTKQR 674
QY 328 GNHTEALEMVTTFGMEDEK-----FSAALVLTSLFKMSKESFERAKRAQSPFLAF 376
DB 675 ---FEEMTNNYQKBIENKKISEGKLGEVEKAKATVDEAVKLOKEIDLRCQHKIAEWAL 731
QY 377 KEAAATKQ-----LAVLSVMQCMETHKLDPAKELPGWQIKQIEIVSLEKDTLQD 425
DB 732 MEKHGHQYDKIVEERDSSELGLYKNRECEQSSAKIALETLS--NIRNELVSLKK---QLE 786
QY 426 KEMBEKARSLSIMEBAALAKRMWQIKRPLSPMPPVTSSSYSPYDRDSFPQR--483
DB 787 IEKEEKEKLKMAKENTAILK---DKDKTKQASLLESPEATSKWFD---SKTTPSQNIS 839
QY 484 -----DDQDETALVSSYLGPSSTFFPHRRRSPEYVMVPLPHGGLGRSVYAYEH 532
DB 840 RLSSSDMSGSKDNDRNLASAKSIL-PTTV-----TKEYVTKTP---TKKSIQRE- 887
QY 533 LAPNSYSPGHGRLHRQ 549
DB 888 ---NKYIPTGGSNKKRK 901

SCPI RAT
ID -SCPI RAT STANDARD; PRT; 997 AA.
AC Q03410;
DT 01-JUN-1994 (Rel. 29, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Synaptonemal complex protein 1 (SCP-1 protein).
GN SYCP1 OR SCPI.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RX MEDLINE=93099884; PubMed=1464329;
RA Meuwissen R.L.J., Offenbergh H.H., Dietrich A.J., Riesenwijk A.,
RA Iersel M., Heyting C.,
RT "A coiled-coil related protein specific for synapsed regions of
RT meiotic prophase chromosomes.";
RL EMBO J. 11:5091-5100(1992).
CC -1- FUNCTION: MAJOR COMPONENT OF THE TRANSVERSE FILAMENTS OF
CC SYNAPTONEMAL COMPLEXES (SCS), FORMED BETWEEN HOMOLOGOUS
CC CHROMOSOMES DURING MEIOTIC PROPHASE.
CC -1- SUBCELLULAR LOCATION: NUCLEAR. IN TRIPARTITE SEGMENTS OF
CC SYNAPTONEMAL COMPLEXES, BETWEEN LATERAL ELEMENTS IN THE NUCLEUS.
CC FOUND ONLY WHERE THE CHROMOSOME CORES ARE SYNAPSED. ITS N-TERMINUS
CC IS FOUND TOWARDS THE CENTRE OF THE SYNAPTONEMAL COMPLEX WHILE THE
CC C-TERMINUS EXTENDS WELL INTO THE LATERAL DOMAIN OF THE
CC SYNAPTONEMAL COMPLEX.
CC -1- TISSUE SPECIFICITY: Testis.
CC -1- DEVELOPMENTAL STAGE: EXPRESSED EXCLUSIVELY IN MEIOTIC PROPHASE
CC CELLS.
CC -1- DOMAIN: CONSISTS OF AN ALPHA-HELICAL STRETCH OF 700 AA RESIDUES,
CC FLANKED BY N- AND C-TERMINAL GLOBULAR DOMAINS. THE C-TERMINAL
CC DOMAIN HAS DNA-BINDING CAPACITY.
CC -1- CAUTION: THIS IS A CONCEPTUAL TRANSLATION; A FRAMESHIFT WAS
CC CORRECTED IN POSITION 6 TO MAXIMIZE THE SIMILARITY WITH THE
CC OTHER SPECIES SYCP1 SEQUENCES.
CC
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; X67805; CA48006.1; ALT FRAME.
KW Nuclear protein; Meiosis; Cell division; Phosphorylation;
KW DNA-binding; Coiled coil.
FT DOMAIN 108 819 COILED COIL (POTENTIAL).
FT DOMAIN 118 121 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DOMAIN 701 704 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DOMAIN 902 905 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DOMAIN 982 990 ARG/LYS-RICH (BASIC).
SQ SEQUENCE 997 AA; 116511 MW; 229D59823FD684BE CRC64;
Query Match 4.2%; Score 129; DB 1; Length 997;
Best Local Similarity 19.1%; Pred. No. 2.3; Indels 146; Gaps 25;
Matches 115; Conservative 106; Mismatches 235;
13 TTPTANPLQRHQSEQRRLPKIVETESTMDITIGSQKQFQFLKSIDELAAAF----SVA 68
385 TCTLEELL---RTEQQRLE-----NNEQDLKITMELQKKS---SELEMTKFKNNKEVE 433
69 VETFK-----RQFDLQKHISIEAIDSKLESNGVLAARNNNFHQPMLSPPRNVSU 122
434 LEEKLTILADQKLLDKKQVEKLAELQKQKQELTFLLQTRKEIH-----DL 482
123 ETTVTVSQPSQE----IVPETSNNKPGGGM---CELMCSKGLFKYIYANISDQAKIMEIP 176
483 EVQVTVTKTSEEHYLQVQEMKTELEKXKNIELTANSMDL-----LLENKKLVQAS 536

177 SALKLAKPEAKFVLDCIGFYIQQBRAFTKESPMSSARQVSLILLESFLLMPDRGKGVK 236
537 DMVLELKHQEDIIINCKQBERMLKQITLEEKWNLDELESVRKEFIQQGDEVKCKL- 595
237 IBSWIKDEAETAAVARWRKRLMTEGGGLAAAARODARGULLLVACFGVPSNF--RSTDLIDL 294
596 -----DKSEENARSIEVEVL-----KKEQMKILENCKNNLKQIENKSNIEEL 640
295 -----IRMSGSN-----IAGALKESQFLVPMVSGIVSESSIKRGVHI 331
641 HOENKALKKSSAENKQNLNAYEIKVKNKLELASTKQKFEEMI-----NNTQKEIEI 692
332 EALEMVTYTFMEDKFSALVLTSLFKMSKESFERAKRAQSPFLAFKEAATKQ----- 383
693 KKISEBKLLGEVEK--AKATVDEAVKLQKEIDLRCOHKIAEWALMEKHQYDKIVEER 750
384 ---LAVLSSVMQCMETHKLDPAKELPGHQIKEQIVLSLEKOTLDKEMEEKARLSLMEE 440
751 DELGLYKNRQEQSSAKVALETELS--NIRNELVSLKK--OLEVEKEBEKELKMEQEN 805
441 AALAKRMVNNQIKRPRLSPMEMPPVTSSSYSPYVDRSRFFSPQR-----DDQ 487
806 TAI---LTDKDKKIQASLLESPEATSNKFD-----SKTTPSQNLSRSSMSDSCSKXDR 858
488 DEISALVSSYLGPSTSPFHRSRSPYVPLPHGGLGRSVYAYEHLAPNSYSPGHGRHLH 547
859 DSLRASAKSILSTTV-----TKEYTVKTF---TKKSIYQRE---NKYLP7GGSNKK 903
548 RQ 549
904 RK 905
RESULT 7
KI67 HUMAN
ID KI67 HUMAN STANDARD; PRT; 3256 AA.
AC P46013;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Antigen KI-67.
GN MKI67.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94043435; PubMed=8227122;
RA Schluter C., Duchrow M., Wohlenberg C., Becker M.H.G., Key G.,
RA Riad H.-D., Gardes J.;
RT "The cell proliferation-associated antigen of antibody Ki-67: a very
RT large, ubiquitous nuclear protein with numerous repeated elements,
RT representing a new kind of cell cycle-maintaining proteins.";
RL J. Cell Biol. 123:513-522(1993).
RN [2]
RP SEQUENCE OF 1-31 FROM N.A.
RA Gardes J., (MAR-1997) to the EMBL/GenBank/DBJ databases.
RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: THOUGHT TO BE REQUIRED FOR MAINTAINING CELL
CC PROLIFERATION.
CC -1- SUBCELLULAR LOCATION: NUCLEAR. PREDOMINANTLY LOCALIZED IN THE G1
CC PHASE IN THE PERINUCLEAR REGION, IN THE LATER PHASES IT IS ALSO
CC DETECTED THROUGHOUT THE NUCLEAR INTERIOR, BEING PREDOMINANTLY
CC LOCALIZED IN THE NUCLEAR MATRIX. IN MITOSIS, IT IS PRESENT ON ALL
CC CHROMOSOMES.
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=Long;
CC IsoId=P46013-1; Sequence=Displayed;
CC Name=Short;
CC IsoId=P46013-2; Sequence=VSP_004298;

220	QY	ILSFILMPD-----RGKGVKIESWIKDEASTAA-VAVRKELMTTEGGAAAEKWDAR	271
1066	DB	--RTFKESPKQILDPAARVTGMKKWPRTPKBEASOLEDLAGFKELFOTPGPSESMTDEK	1123
272	QY	GLLLLVACFGVP-----SNFRSTDL---LDLIRMGSNEIAGALKESQF	312
1124	DB	--TTKIACKSPPEVDTPSTKQWPKESLRKADVEBEFLALRLKLTPS---AG---KAM	1174
313	QY	LVPWVGIVYESSIK--RGNHIEALEWVVTFGMEDKFSAAVLITSLFKWSKESPERAKKA	370
1175	DB	LTPAPAGGDEKDKAFMGTPVQKDLAGT-----LPGSKRQLQOTPKKA	1218
371	QY	QS--PLA-FKE-----AATKQ-----	383
1219	DB	QALLEDLAGFKELFOTPGHTEELVAAGKTTKPCDSPQSDPVDTPSTKQRPKRSIRKADV	1278
384	QY	---LA---VLSSWQCMET-----HKLDPAKELPGWQIKKEQIVSL	417
1279	DB	EGELLACRNLPAGKAMHTPKPSVGECKDIIIFVGTVPQKDLTENTLTSKRRPQTPKE	1338
418	QY	EKOTQLQDKEMEEKARSLSMBEAAALAKRMVNNQIKRPLRSLPMEMPVPTSSSYPIYDR	477
1339	DB	EAQALDLTGFKELFOTPGHTEEAAG-----KTTKMPCESSPESADTPTSTRQP	1391
478	QY	SFSPQRDDDDDETSAL--VSSVLGPSTSPPHRSRESPEVMVPLPHGGGLGRSVAYAEHLAP	535
1392	DB	KTPLEKRDVQKELSAUKUTQTSGETT---HTDK-----VP---GGEDKSINAFRETAK	1439
536	QY	NSYSPGHGHLRHQYSPSLVHQORHP	561
1440	DB	-----AASVTGSKRHP	1455

RESULT 8

ID	PCT2_MOUSE	STANDARD;	PRT; 1920 AA.
ACT	PCT2_MOUSE		
ADT	P48725;		
DDT	01-FEB-1996 (Rel. 33, Created)		
DDT	01-FEB-1996 (Rel. 33, Last sequence update)		
DDT	28-FEB-2003 (Rel. 41, Last annotation update)		
DE	Pericentrin 2		
DS	PCT2 OR PCNT		
GN	Mus musculus (Mouse).		
RP	Medline=94170365; PubMed=8124707;		
RR	Dorsey S.J., Stein P., Evans L., Calarco P.D., Kirschner M.;		
RA	"Pericentrin, a highly conserved centrosome protein involved in		
RT	microtubule organization.";		
RL	Cell 76:639-650 (1994).		
RL	Cell 76:639-650 (1994).		
CC	FUNCTION: INTEGRAL COMPONENT OF THE FILAMENTOUS MATRIX OF THE		
CC	CENTROSOME INVOLVED IN THE INITIAL ESTABLISHMENT OF ORGANIZED		
CC	MICROTUBULE ARRAYS IN BOTH MITOSIS AND MEIOSIS.		
CC	TISSUE SPECIFICITY: HIGHEST LEVELS FOUND IN KIDNEY, THYMUS AND		
CC	LIVER. LOW LEVELS IN BRAIN, MUSCLE, LUNG AND HEART.		
CC	DOMAIN: COMPOSED OF A COILED-COIL CENTRAL REGION FLANKED BY NON-		
CC	HELICAL N- AND C-TERMINALS.		
CC	SIMILARITY: STRONG, TO HUMAN KENDRIN.		

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).

EMBL; U05823; AAA17886.1; -
 DR
 PIR; A53188; A53188.

DR MGD; MGI:102722; Pcnt2.
 DR GO; GO:0005813; C:centrosome; IDA.
 KW Coiled coil; Microtubules.
 FT DOMAIN 110 1600
 FT DOMAIN 109 112
 FT DOMAIN 383 387
 FT DOMAIN 1920 AA; 218337 MW; CF1D0ADBC5B73309 CRC64;
 SQ SEQUENCE 1920 AA; 218337 MW; CF1D0ADBC5B73309 CRC64;
 Query Match 4.1%; Score 126.5; DB 1; Length 1920;
 Best Local Similarity 19.8%; Pred. No. 7.8; Indels 141; Gaps 25;
 Matches 105; Conservative 99; Mismatches 186; Indels 141; Gaps 25;
 21 QRHSQRRELPKIVETSTMDITIG---QSKQPOFLKSIDELAAPSVA-VETFKRQF 76
 679 QLASSESRQAL---LETHVAELQVKNHAEISALEKRLHSLNLEDESCYADVQVIR--- 732
 77 DLQKHIEIENAIKSLGNSGVLAARNPHQMLSPRRNV-----SVETVT 127
 733 DEHQALELLRAELBEQLQ-----KXESCHREMLTQELNKRQHAELQSVRSLR 784
 128 VSQPSQEIIV-----PETSINKPEGGRMCE-----LMCSKGLR 158
 785 MENSAGHIENGKGPADLOGAQHQQDPAMALHNEGHLLVEDGDVILRSVDAEGLLHQAGPQ 844
 159 KYIYANISQAKMEI---PSALKIAKEPAKFVLDICIGKFVLOGR-----PAFTKESPM 211
 845 ELGDAHTVEMQKSAELAKPQELQASQDQVQRD---KVFLLNRELECECRADVEQLQQR 901
 212 SARQ---VSLILFELFMPGRGKVKIESWIKD-----EAEATAAVARIKRLM 257
 902 RERENEGTTLICMLRADLELAQGEKA-----LDAALRULLDFGDTLKAATVILKSRIS 956
 258 TEGGLA-----AARKMDARGLLLVACFG-----VPSNFRSTDLIDL 294
 957 EGAGLLLDHEDADTSDAR---LAAAPLGDMMWSDGLLEIDRTLPEGAETSSVCEISHV 1013
 295 -----IRMSGNEIAGALKR-SQFLVPMVSGIVESIKRGHIEALEMVTYFGMEDKPSA 348
 1014 CSEFFRPNENTLDCQPIRVVQSLSSTAVEGLEHVALDSSKOLEARQLHRC-VEREF-- 1070
 349 ALVITSFLKMSKESFBRARAKQSPPL-----AFKEAATKOLAVLSSVMQCMETHKLD 401
 1071 -----RHRNEBMAQAMQKQELLERLREBSAARDLALHTAKGLLEGFKVEKYVDL 1122
 402 AKELPGWQIKQEVLSLEKTLQDKMEMEKARS-LSLMEEAALAKEMYNQO 451
 1123 QEALGKKKESEQQILLEBDR--KOLEQAARELTLKEEKSV---LWNQK 1168
 RESULT 9
 SR72_CAEL STANDARD; PRT; 694 AA.
 ID P91240;
 AC 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Signal recognition particle 72 kDa protein homolog (SRP72).
 GN F08D12.1
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoides;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN SEQUENCE FROM N2.
 RC STRAIN=Bristol N2.
 EC Le T.T., Waterston R.;
 RA Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
 RL FUNCTION: SIGNAL-RECOGNITION-PARTICLE ASSEMBLY HAS A CRUCIAL ROLE
 CC IN TARGETING SECRETORY PROTEINS TO THE ROUGH ENDOPLASMIC
 CC RETICULUM MEMBRANE. SRP72 BINDS THE 7S RNA ONLY IN PRESENCE OF
 CC SRP68. THIS RIBONUCLEOPROTEIN COMPLEX MIGHT INTERACT DIRECTLY WITH
 CC THE DOCKING PROTEIN IN THE ER MEMBRANE AND POSSIBLY PARTICIPATE
 CC IN THE ELONGATION ARREST FUNCTION (BY SIMILARITY).

CC -!- SUBUNIT: SIGNAL RECOGNITION PARTICLE CONSISTS OF A 7S RNA MOLECULE
 CC OF 300 NUCLEOTIDES AND SIX PROTEIN SUBUNITS: SRP72, SRP68, SRP54,
 CC SRP19, SRP14 AND SRP9 (BY SIMILARITY).
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -!- SIMILARITY: BELONGS TO THE SRP72 FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; U80840; AAB37925.1; -;
 CC FIR; T25685; T25685.
 CC Wormpep; F08D12.1; CE09234.
 CC InterPro; IPR001440; TPR.
 CC Pfam; PF00515; TPR; 3
 KW Signal recognition particle; Ribonucleoprotein.
 FT DOMAIN 684 689
 FT SEQUENCE 694 AA; 78485 MW; E512B69DC8B7C4FA CRC64;
 SQ SEQUENCE 694 AA; 78485 MW; E512B69DC8B7C4FA CRC64;
 Query Match 4.1%; Score 126; DB 1; Length 694;
 Best Local Similarity 18.9%; Pred. No. 2.1;
 Matches 111; Conservative 92; Mismatches 199; Indels 186; Gaps 24;
 35 IYETESTND-----ITIGOSQPOFLKSIDELAAPSVAVETFKRQFDLQKHISIE 88
 164 VAETEDSYSLYNACVIEAEKLPQALESEK-----ALKTKSFEDEREDEIE 217
 89 ATDKSLGNSGVVLAARNN---FHQPMSPRRNVSVETTVTSQP---SQEIVPETS 140
 218 EUDSTRVKAYVQLRMGQKAEALAIYKVAANHPDSVKATITNNIPASSDFALPESR 277
 141 NKPEGG-RMCELMCSK-----GLRKYIYANISDQAKMEIIPSA 178
 278 KRFAALQIFQSKCSKPPDFWLKMLQSKSKTKFLQKNGNFTYFTSKNSQIYRFSQAIPPK 337
 179 LKLAKEPAKFVLDICIGKFYLOGRRAFTKESPSWSARQVSLILLESFLIMPDRGKVKIE 238
 338 -----IPNFSFO-----IDTKLTRRQRLTMLNNALVLL 367
 239 SWIKDEAETAAVAMKRLMTGGLAAAEKMDARGLLLVACFG-----VPSNFRST 289
 368 --LSNQREPC-----KRALEE-----LVAKFGSSKDVALTEATLHFQMG 404
 290 DLDLIRMSGNEIAGALKRSQFLVPMVSGIVESIKRGHIEALEMVTYFGMEDKPSAA 349
 405 DAEAAKVLGSDLEQSLARHLV-----LNAGRLPEAVGAIRDLFISGKLGAS 453
 350 LVLTSTFL-----KMSKESPERAKRKAQSPAPFAKE-----AATKOLAV 386
 454 SULTSTLIAADSRDEAVKELVAASTAKNQTPALKSILEDLVEVEQQGNETATKHEK 513
 387 L-----SSVMQCM---ETHKLDPAKELPGWQIKQEVLSLEKDTLQDKEMEE 430
 514 LVEKFPEDLQQLRLVGAYSKTDPKK-----AESLSAKLFPETMEVDVNVNDELESDWI 567
 431 -----KARSLMEEAALAKEMYNQOIKRPLSPMEMPPVTSSVSP-----IY 474
 568 LYGEKTRQKKEAKSPQTAETATRK-----LKATKRRKIRLPKNYNSAVTPDPERMLPR 623
 475 RDRS-FPSQORDDDQDEI--SALVSSYLGPSTSFPHRSRRSPPEYVMVPLP 519
 624 QERSTYKRRKKNREIGRGTCGSSSANPNVEYVTASPNSPR---PLP 668
 RESULT 10
 GOG4_HUMAN
 ID GOG4_HUMAN
 AC Q13439; Q13270; Q13654; Q14436;
 DT 16-OCT-2001 (Rel. 40, Created)

16-OCT-2001 (Rel. 40, Last sequence update)
15-SEP-2003 (Rel. 42, Last annotation update)
Golgi autotransfer, golgin subfamily A 4 (Trans-Golgi p230) (256 kDa golgin) (Golgin-245) (72.1 protein).
GOLGA4.
Homo sapiens (Human).
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCBI_TaxID=9606;
[1]
SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
MEDLINE=96215236; PubMed=8626529;
Erich R., Gleeson P.A., Campbell P., Dietzsch E., Toh B.-H.,
"Molecular characterization of trans-Golgi p230: a human peripheral membrane protein encoded by a gene on chromosome 6p12-22 contains extensive coiled-coil alpha-helical domains and a granin motif.";
J. Biol. Chem. 271:8328-8337(1996).
[2]
SEQUENCE FROM N.A.
Seelig H.P.;
Submitted (NOV-1994) to the EMBL/GenBank/DBJ databases.
[3]
SEQUENCE OF 131-2230 FROM N.A.
TISSUE=Placenta;
MEDLINE=96125112; PubMed=8537393;
Fritzler M.J., Lung C.-C., Hamel J.C., Griffith X.J., Chan E.K.L.;
"Molecular characterization of golgin-245, a novel Golgi complex protein containing a granin signature.";
J. Biol. Chem. 270:31262-31268(1995).
[4]
SEQUENCE OF 524-672 FROM N.A.
TISSUE=Gastric fundus;
Balague C.;
Thesis (1994), Instituto municipal de investigacion medica, Spain.
-!- FUNCTION: May play a role in vesicular transport from the trans-Golgi.
-!- SUBCELLULAR LOCATION: ASSOCIATED WITH THE CYTOPLASMIC FACE OF THE GOLGI MEMBRANE.
-!- ALTERNATIVE PRODUCTS:
Event=Alternative splicing; Named isoforms=4;
Comment=Additional isoforms seem to exist;
Name=1;
IsoId=Q13439-1; Sequence=Displayed;
Name=2;
IsoId=Q13439-2; Sequence=VSP_004272, VSP_004273;
Name=3;
IsoId=Q13439-3; Sequence=VSP_004274;
Name=4;
IsoId=Q13439-4; Sequence=VSP_004275;
-!- DISEASE: ANTIGEN IN THE AUTOIMMUNE DISEASE SJOEGREN'S SYNDROME AND IN HEPATITIS B.
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
EMBL; U41740; AAC50434.1; -;
EMBL; X82834; AAC58041.1; -;
EMBL; U31906; AAC51791.1; -;
EMBL; X76942; AAC54261.1; -;
Genew; HGNC:4427; GOLGA4.
MIM; 602509; -;
GO; GO:0005802; C:Golgi trans-face; TAS.
GO; GO:0016192; P:vesicle-mediated transport; TAS.
InterPro; IPR000237; GRIP_domain.
Pfam; PF01465; GRIP; 1.
Golgi stack; Antigen; Coiled coil; Alternative splicing.
COILED COIL (POTENTIAL).
DOMAIN 133 237
DOMAIN 276 1011
COILED COIL (POTENTIAL).

FT DOMAIN 1033 1214 COILED COIL (POTENTIAL).
FT DOMAIN 1259 2152 COILED COIL (POTENTIAL).
FT VARSPDIC 2154 2185 TPYKGNLHTDVSLFGPTFEYLRKVLFEY -> HLTKV
AICTRMSSHLNPLNLSICKCKFLSI (in isoform 2).
FT FT /FTId=VSP_004272.
FT FT Missing (in isoform 2).
FT VARSPDIC 2186 2230 /FTId=VSP_004273.
FT FT Missing (in isoform 3).
FT VARSPDIC 2103 2109 /FTId=VSP_004274.
FT VARSPDIC 2222 2230 /FTId=VSP_004275.
FT FT R -> K (in REF. 3).
FT CONFLICT 188 188 Y -> H (in REF. 3).
FT CONFLICT 220 220 T -> A (in REF. 3).
FT CONFLICT 276 276 K -> E (in REF. 3).
FT CONFLICT 584 584 T -> A (in REF. 3).
FT CONFLICT 628 628 K -> E (in REF. 3).
FT CONFLICT 630 630 K -> N (in REF. 3).
FT CONFLICT 682 682 K -> N (in REF. 3).
SQ SEQUENCE 2230 AA; 261139 MW; 3BBY733DB1EA86134 CRC64;
Query Match 4.1%; Score 126; DB 1; Length 2230;
Best Local Similarity 20.9%; Pred. No. 10;
Matches 120; Conservative 79; Mismatches 204; Indels 170; Gaps 28;
QY 3 NYPTVAAQPTTANPLLRQHS----EORRELPK-IVETESTSMDITIGSKQPOFLK 57
DB 805 SYGSATHEQTKAYEEQLAQLOKLLDLETERILLTKQVAEVAQKDV----- 852
QY 58 SIDELAAFSVAVETFKROF---DLQKHIEINAIDSKLESNG-----VVLARNN 106
DB 853 -CTELDAHKIQVDLMQOEKQNSEMEQVKSLTQVYESKLEDGKEQEQTKQILVEKEN 911
QY 107 NFHPMLSPPRNNVSVET-TVTVSPQSOBIVP---ETSNKPEGSMEL-MCSKGLRKVI 161
DB 912 MILQ-MREGGKKEIILTKLAKEDSIHILNEEYETFKQEKQKVKQKAKEMOETL 970
QY 162 YANISQ-AKLMEIIP-SALKLAKPAKEVLDCIGKFLYQGRRAFTKESPMSSARQVSL 219
DB 971 KKLLQOEAKLKELENTALELSQKEQP-----NAKMLEQAANSAGIS 1015
QY 220 ILBSFLIMPDRGKGVKIBSWI---KDEATAVAWRKELMTEGGLAAAEKM----- 268
DB 1016 DAVSRL-----ETNQKEIESLTVHRRLNDVISIWEKLNQ-----AEELQEIHEIQL 1066
QY 269 -----DARGLLLVACGVSPNFRSTDLLD--LIRMSGSEIAGALKRSQFLVPMVS 318
DB 1067 QEKEQVAVELKQKILLFGCEKEEMKEITWLKEEGVKQDTTLNELOEQLOKQS----AHVN 1123
QY 319 GIVESSIKRGMHTEALEM-----VYTFGMEDKFSALVLTSLRMSK 360
DB 1124 SLAQDETKLKAHLEKLEVDINKSLKENTFLQBLQVELKMLAEEDRKVS-ELTSKLKTTD 1182
QY 361 ESTERAK-----RKAQSLAPK-----BAATQQL----- 384
DB 1183 EEFQSLKSSHEKSNKSLDKSLFKSLSELAIQDICKCKTEALLEAKTNELINISSK 1242
QY 385 --AVLSSVMOCMETHKLDPAKELPGQIKQIVS-LEKDTLQLDKE----- 427
DB 1243 TNALSRISHCQ--HRTTKVKE--ALLIKTIVSELAQLRQLTBEQNTLISFOQATHQ 1298
QY 428 MEKARSLSLM-----EEAALAKRMYNQ 451
DB 1299 LEEKENQIKSMKADIESLVTEKEALQKEGNNQ 1331
RESULT 11
SNE2_HUMAN
ID SNE2_HUMAN STANDARD; PRT; 6885 AA.
AC Q8WXH0; Q8N1S3; Q8NF49; Q8TER7; Q8WNW3; Q8WNW4; Q8WNW5; Q8WXH1;
AC Q9NU50; Q9UF04; Q9Y2L4; Q9Y4R1;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)

DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Nesprin 2 (Nuclear envelope spectrin repeat protein 2) (Syne-2)
 DE (Synaptic nuclear envelope protein 2) (Nucleus and actin connecting
 DE element protein) (NUANCE protein).
 GN SYNE2 OR NUA OR KIAA1011.
 OS Homo sapiens (Human)
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN
 RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 9), FUNCTION, CHARACTERIZATION, AND
 RP INTERACTION WITH F-ACTIN.
 RX MEDLINE=22113122; PubMed=12118075;
 RA Zhen Y.-Y., Libotte T., Munk M., Noegel A.A., Korenbaum E.;
 RT "NUANCE, a giant protein connecting the nucleus and actin
 RT cytoskeleton."; *J. Cell Sci.* 115:3207-3222(2002).
 RL [2]
 RN SEQUENCE FROM N.A. (ISOFORMS 4; 5 AND 7).
 RX MEDLINE=21652858; PubMed=11792814;
 RA Zhang Q., Skepper J.N., Yang F., Davies J.D., Hegyi L., Roberts R.G.,
 RA Weissberg P.L., Ellis J.A., Shanahan C.M.;
 RT "Nesprins: a novel family of spectrin-repeat-containing proteins that
 RT localize to the nuclear membrane in multiple tissues."; *J. Cell Sci.* 114:4485-4498(2001).
 RL [3]
 RN SEQUENCE FROM N.A. (ISOFORM 1).
 RP MEDLINE=22296983; PubMed=12408964;
 RX Zhang Q., Ragnauth C., Greener M.J., Shanahan C.M., Roberts R.G.;
 RT "The nesprins are giant actin-binding proteins, orthologous to
 RT Drosophila melanogaster muscle MSP-300."; *Genomics* 80:473-481(2002).
 RL [4]
 RN SEQUENCE FROM N.A. (ISOFORM 6).
 RP TISSUE=Testis;
 RC Poustka A., Klein M., Mewes H.-W., Gassenhuber J., Wiemann S.;
 RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RN SEQUENCE FROM N.A.
 RP PubMed=12508121;
 RX Heilig R., Eckenberg R., Petit J.-L., Fonknechten N., Da Silva C.,
 RA Cattolico L., Levy M., Barbe V., De Berardinis V., Ureta-Vidal A.,
 RA Pelletier E., Vico V., Anthouard V., Rowen L., Madan A., Qin S.,
 RA Sun H., Du H., Pepin K., Artiguenave F., Robert C., Craud C.,
 RA Brule T., Jaillon O., Friedlander L., Samson G., Brottier P.,
 RA Cure S., Segurens B., Aniers F., Samain S., Crepeau H., Abbasi N.,
 RA Alach N., Boscus D., Dickhoff R., Doris M., Dubois I., Friedman C.,
 RA Gouyvenoux M., James R., Madan A., Maltrey-Estrada B., Mangenot S.,
 RA Martins N., Menard M., Oztas S., Ratcliffe A., Shaffer T., Traek B.,
 RA Vacherie B., Bellemere C., Belser C., Besnard-Gonnet M.,
 RA Bartol-Mavel D., Boutard M., Briez-Silla S., Combette S.,
 RA Dufosse-Laurent V., Ferron C., Lechaplais C., Louesse C., Musset D.,
 RA Magdelenat G., Pateau E., Petit E., Sirvain-Trukniewicz P., Trybou A.,
 RA Vega-Czarny N., Bataille E., Bluet E., Bordenais I., Dubois M.,
 RA Dumont C., Guerin T., Hafray S., Hamadi R., Munga J., Pellouin V.,
 RA Robert D., Wunderle E., Gauguet G., Roy A., Sainte-Marthe L.,
 RA Verdier J., Verdier-Discala C., Hillier L., Fulton L., McPherson J.,
 RA Matsuda F., Wilson R., Scarpetti C., Gyapay G., Wincker P., Saurin W.,
 RA Quetier F., Waterston R., Hood L., Weissbach J.;
 RT "The DNA sequence and analysis of human chromosome 14."; *Nature* 421:601-607(2003).
 RL [6]
 RN SEQUENCE FROM N.A. (ISOFORM 8).
 RP TISSUE=Brain;
 RC MEDLINE=22398257; PubMed=12477932;
 RX Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Schuler G.D.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buettow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.E., Scheetz T.E.,
 RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahy J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalek U., Smallus D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences."; *Proc. Natl. Acad. Sci. U.S.A.* 99:16899-16903(2002).
 RL [7]
 RN SEQUENCE FROM N.A. (ISOFORM 3), AND SEQUENCE OF 1-956 AND 5133-6885
 RP FROM N.A.
 RC TISSUE=Spleen, and Tongue;
 RA Jikuya H., Takano J., Nomura N., Kikuno R., Nagase T., Ohara O.,
 RA Ninomiya K., Wagatsuma M., Kanda K., Kondo H., Yokoi T., Kodaira H.,
 RA Furuya T., Takahashi M., Kikawa E., Omura Y., Abe K., Kamihara K.,
 RA Katsuta N., Sato K., Tanikawa M., Yamazaki M., Sugiyama T., Irie R.,
 RA Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,
 RA Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,
 RA Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Murakawa K.,
 RA Kanehori K., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B.,
 RA Suzuki Y., Sugano S., Nagahari K., Masuho Y., Nagai K., Isogai T.;
 RT "NEO human cDNA sequencing project."; Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
 RL [8]
 RN SEQUENCE OF 5709-6885 FROM N.A. (ISOFORM 2).
 RP TISSUE=Brain;
 RX MEDLINE=99246063; PubMed=10231032;
 RA Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Hitosawa M.,
 RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes. XIII.
 RT The complete sequences of 100 new cDNA clones from brain which code
 RT for large proteins in vitro."; *DNA Res.* 6:63-70(1999).
 RL [9]
 RN REVISIONS.
 RP TISSUE=Brain;
 RX MEDLINE=22158633; PubMed=12168954;
 RA Nakajima D., Okazaki N., Yamakawa H., Kikuno R., Ohara O., Nagase T.;
 RT "Construction of expression-ready cDNA clones for KIAA genes: manual
 RT curation of 330 KIAA cDNA clones."; *DNA Res.* 9:99-106(2002).
 RL [10]
 RN SEQUENCE OF 5754-6885 FROM N.A.
 RP MEDLINE=21154917; PubMed=11230166;
 RX Wiemann S., Weil B., Wellenreuther R., Gassenhuber J., Glassl S.,
 RA Ansorge W., Boecker M., Bloeker H., Bauersachs S., Blum H.,
 RA Lauber J., Duesterhoeft A., Beyer A., Koehler K., Strack N.,
 RA Mewes H.-W., Othenwaelder B., Obermaier B., Tampe J., Heubner D.,
 RA Wambutt R., Korn B., Klein M., Poustka A.;
 RT "Towards a catalog of human genes and proteins: sequencing and
 RT analysis of 500 novel complete protein coding human cDNAs."; *Genome Res.* 11:422-435(2001).
 RL CC
 CC -1- FUNCTION: Involved in the maintenance of nuclear organization and
 CC structural integrity. Probable anchoring protein which thwarts the
 CC nucleus to the cytoskeleton. Connects nuclei to the cytoskeleton
 CC by interacting with the nuclear envelope and with F-actin in the
 CC cytoplasm.
 CC
 CC -1- SUBUNIT: Interacts with F-actin via its N-terminal domain.
 CC
 CC -1- SUBCELLULAR LOCATION: Type IV membrane protein (Potential). The
 CC largest part of the protein is cytoplasmic, while its C-terminal
 CC part is associated with the nuclear envelope, most probably the
 CC outer nuclear membrane. Remains associated with the nuclear
 CC envelope during its breakdown in mitotic cells.
 CC
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event-Alternative splicing; Named isoforms=9;
 CC Name=1;
 CC IsoId=Q8WXH0-1; Sequence=Displayed;
 CC Name=2;
 CC IsoId=Q8WXH0-2; Sequence=VSP_007164, VSP_007166;
 CC

CC Notes=No experimental confirmation available;
 CC Name=3;
 CC IsoId=Q8WXH0-3; Sequence=VSP_007155;
 CC Note=Produced by exon skipping that results in a frameshift. No
 CC experimental confirmation available;
 CC Name=4; Synonym=Beta;
 CC IsoId=Q8WXH0-4; Sequence=VSP_007156;
 CC Name=5; Synonym=Alpha;
 CC IsoId=Q8WXH0-5; Sequence=VSP_007157, VSP_007164, VSP_007165;
 CC Name=6;
 CC IsoId=Q8WXH0-6; Sequence=VSP_007158, VSP_007165, VSP_007166;
 CC Note=No experimental confirmation available;
 CC Name=7; Synonym=Gamma;
 CC IsoId=Q8WXH0-7; Sequence=VSP_007154, VSP_007163;
 CC Name=8;
 CC IsoId=Q8WXH0-8; Sequence=VSP_007161, VSP_007162;
 CC Note=No experimental confirmation available;
 CC Name=9; Synonym=Nuance-N-33;
 CC IsoId=Q8WXH0-9; Sequence=VSP_007159, VSP_007160;
 CC TISSUE SPECIFICITY: Widely expressed with higher level in kidney,
 CC adult and fetal liver, stomach and placenta. Weakly expressed in
 CC skeletal muscle and brain. Isoform 5 is highly expressed in
 CC pancreas, skeletal muscle and heart.
 CC DOMAIN: The Klarsicht domain mediates the nuclear envelope
 CC targeting.
 CC SIMILARITY: Belongs to the Nesprin family.
 CC SIMILARITY: Contains 1 actin-binding domain.
 CC SIMILARITY: Contains 2 calponin-homology (CH) domains.
 CC SIMILARITY: Contains 1 Klarsicht domain.
 CC SIMILARITY: Contains 13 leucine-rich (LRR) repeats.
 CC SIMILARITY: Contains 9 spectrin repeats.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AF435010; AAL33547.1; -
 CC EMBL; AF435011; AAL33548.1; -
 CC EMBL; AY061757; AAL33800.1; -
 CC EMBL; AY061758; AAL33801.1; -
 CC EMBL; AY061759; AAL33802.1; -
 CC EMBL; AF495911; AANG0443.1; -
 CC EMBL; AL117404; CAB55905.1; -
 CC EMBL; AL162832; -; NOT ANNOTATED CDS.
 CC EMBL; AL355094; -; NOT ANNOTATED_CDS.
 CC
 CC Query Match 4.0%; Score 125.5; DB 1; Length 6985;
 CC Best Local Similarity 18.9%; Pred. No. 50;
 CC Matches 100; Conservative 89; Mismatches 190; Indels 149; Gaps 20;
 CC
 CC 57 KSIDLAAPSVAVETFKQFDLQ-----KHISIEAIDSKL 94
 CC 3214 KORENSEADVETKLEFDELQVLTNTVDLNTVNDAYENLRYKAVRAVESIT 3273
 CC
 CC 95 ESNQGVLAARN--NFHOMLSPRRNVSVETVY-----VSQPS----- 132
 CC 3274 SLEAIIIVYDVGNPESEMLPKQBELESTVAHQDLTEKLGMTSSPEAKLQVYTL 3333
 CC
 CC 133 QEVIVETSKPEGGMCELMCKGLRY-----IYANISDQAKLMEIEPSALKLAK 183
 CC 3334 QELVSKNKAEMKFAQTEAEYRLENKYCKRMKEEDYINLSKMETVIGQSSSPLSY 3393
 CC
 CC 184 EPAKFLVDCIGKFLQGRRAFTKSPSSAQQVLSFLMLPDRGKGKVI-----E 238
 CC 3394 REALERLE--QSKALVSNLSITKBEELMKLRQILRLR---CTENDGICLLKIVSALWE 3448
 CC
 CC 239 SWIK--DEAETAAVAKRKLMTGGGLAAEKVDAKGLILLVACGVSPNFS----- 288
 CC 3449 KWSLSLEAKENWEMCEELKQWKF-VSEIEIERAIIIDNLQBELPEISKTEAATTEEL 3507

RESULT 12

BEM3_YEAST

ID BEM3_YEAST STANDARD; PRT: 1128 AA.

AC P32873;

DT 01-OCT-1993 (Rel. 27, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE GTPase-activating protein BEM3.

GN BEM3 OR YPL115C OR LPH12C.

OS Saccharomyces cerevisiae (Baker's yeast).

OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.

OX NCBI_TaxID=4932;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=94043316; PubMed=8227021;

RA Zheng Y., Hart M.J., Shinjo K., Evans T., Bender A., Cerione R.A.;

RT "Biochemical comparisons of the Saccharomyces cerevisiae Bem2 and

RT Bem3 proteins. Delineation of a limit Cdc42 GTPase-activating protein

RT domain.";

RL J. Biol. Chem. 268:24629-24634(1993).

RN [2]

RP SEQUENCE FROM N.A.

RX STRAIN=S288C / AB972;

RX MEDLINE=97313271; PubMed=9169875;

RA Bussey H., Storms R.K., Ahmed A., Albermann K., Allen E., Ansoorge W.,

RA Araujo R., Aparicio A., Barrell B.G., Badcock K., Benes V.,

RA Botstein D., Bowman S., Bruckner M., Carpenter J., Cherry J.M.,

RA Chung E., Churcher C.M., Coster F., Davis K., Davis R.W.,

RA Dietrich F.S., Delius H., DiPaolo T., Dubois E., Duesterhoeft A.,

RA Durcan M., Floeth M., Fortin N., Friesen J.D., Fritz C., Goffeau A.,

RA Hall J., Hebling U., Heumann K., Hilbert H., Hillier L.,

RA Hunnicke-Smith S., Hyman R., Johnston M., Kalman S., Kleine K.,

RA Komp C., Kurdi O., Lashkari D., Lew H., Lin A., Lin D., Louis E.J.,

RA Marathe R., Messenguy F., Mewes H.-W., Mirtipati S., Moestl D.,

RA Mueller-Auer S., Namath A., Newtich U., Oefner P., Pearson D.,

RA Petel F.X., Pohl T.M., Purnelle D., Schafer M., Scharfe M.,

RA Scherens B., Schramm S., Schroeder M., Sidcu A.M., Tettelin H.,

RA Urrutara L.A., Ushinsky S., Vierendeels F., Vissers S., Voss H.,

RA Walsh S.V., Wambutt R., Wang Y., Wedler E., Wedler H., Winnett E.,

RA Zhong W.W., Zollner A., Vo D.H., Hani J.;

RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome XVI.";

RL Nature 387:103-105(1997).

CC -!- FUNCTION: GTPASE-ACTIVATING PROTEIN (GAP) FOR CDC42 AND LESS

CC EFFICIENTLY FOR RHO1 NEGATIVE REGULATOR OF THE PHEROMONE-RESPONSE

CC PATHWAY THROUGH THE STE20 PROTEIN KINASE.

CC -!- SUBCELLULAR LOCATION: Cytoplasmic.

CC -!- SIMILARITY: TO REGULATORY DOMAIN OF PROTEIN KINASE C AND PCR,

CC THE PRODUCT OF THE BREAKPOINT CLUSTER REGION GENE.

CC -!- SIMILARITY: Contains 1 PH domain.

CC -!- SIMILARITY: Contains 1 rho-GAP domain.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC ENBL; L14558; AAB34453.1; -;
 DR ENBL; U43503; AAB68247.1; -;
 DR PIR; A49960; A49960.
 DR SGD; S0006036; BEM3.
 DR GO; GO:0005100; F:rho GTPase activator activity; IDA.
 DR GO; GO:0004871; F:signal transducer activity; IPI.
 DR GO; GO:0007118; P:apical bud growth; IPI.
 DR GO; GO:0000283; P:establishment of cell polarity (sensu Sacch. . .; IPI.
 DR GO; GO:000125; P:invasive growth; IPI.
 DR GO; GO:0007119; P:isotropic bud growth; IPI.
 DR GO; GO:0007124; P:pseudohyphal growth; IPI.
 DR GO; GO:0007264; P:small GTPase mediated signal transduction; IPI.
 DR InterPro; IPR001849; PH.
 DR InterPro; IPR001683; PX.
 DR InterPro; IPR000198; RhogAP.
 DR Pfam; PF00169; PH; 1.
 DR Pfam; PF00787; FX; 1.
 DR Pfam; PF00620; RhogAP; 1.
 DR SMART; SM00233; PH; 1.
 DR SMART; SM00312; PX; 1.
 DR SMART; SM00324; RhogAP; 1.
 DR PROSITE; PS50003; PH DOMAIN; 1.
 DR PROSITE; PS50238; RhogAP; 1.
 KW GTPase activation.
 FT DOMAIN 634 741
 FT DOMAIN 913 1128
 FT DOMAIN 913 1128
 SQ SEQUENCE 1128 AA; 124912 MW; 4BF03EAD6EF10283 CRC64;

Query Match 4.0%; Score 125; DB 1; Length 1128;
 Best Local Similarity 19.3%; Pred. No. 4.7; Indels 220; Gaps 30;
 Matches 134; Conservative 97; Mismatches 243;

QY 13 TTTANLLQHQSEQRRLPRLKIVETESTMDITIGSKQKQF-----LKSIDEAAAF 65
 DB 12 STTLELAQVNDHRSKKDKSIEHKGTC-----GKERNFSYDEIFENIKLQVQY 66
 QY 66 SVAVETFKQFDLDQV-----IESIEN-AIDSKLESNGVLAARN-----N 107
 DB 67 ETEIESLEKVIDMLQKREASLEVLEQVQNDSDRSYVNDQSFVLPFPAERKAHISLN 126
 QY 108 FHQPMLSPP---RNVSVETTTVTSQPSOEIVPETS-NKPEGGRM-----CELMCSKGLR 158
 DB 127 LPIPTLSPLQOGSDVALETSVTPVQIGVTSNTSISRKLQNLNMLNDEIANSFSSP 186
 QY 159 KIYANISQAKLME-----IPSAKLAKPAKFLV----- 190
 DB 187 KIINRSVSSPTKIHSEQLASPAASVYTTTSTRITIKSPNKGSKPLQERLRSFQNPMTA 246
 QY 191 -----DCICK-----FVLOGRRAFTKESPMSSARQ 215
 DB 247 VINNHLSPLKASTNNLDLTSKSOQLTNDIAIQNDRVSSITSSAYTTGTTSAKS 306
 QY 216 VSLIILEFLMPDRGKGKVKIESWIKDEAETAAVAMVKRLMTEGGLAAAEKMDARGLL 275
 DB 307 PSSLL-----EVK-----EGENKALGFSP-----ASKELDDFTOLL 338
 QY 276 LVACFGVPNSRSTDLDL-----IRMSGNETAGALKRSQFLVPMVSGIVESSIKRG 328
 DB 339 -----DSFGEDLVNTDSKPLSISTINESUPPPAPTFFSPPTSSGINKNSTPLS 391
 QY 329 MHTALEMYVTFGMEDKFAAL-----VLTSFLKMSKESFERAKRKAQ---SPLAFKE 378
 DB 392 SHLAS--PVLNKKDDNFGAQAQAKNKKKRVLTSSLPNLSTKLSTTSQNASLPPNPPESS 449
 QY 379 AATKQLAVLSVMQCMETHKLPAPKELPCWQIKQVSLKDTQLDKMEKARSLM 438
 DB 450 SKOKQLGETASI---HSTNTLNTFTSTPGSLK-----TLR-----RPHASSVSTV 492

QY 439 EBAALAKMYNQIKRP---RLSPWMPPTVSSSYPIYDRSPFSPQDDDDQDEISALVS 495
 DB 493 KSAV-----QSLKSDIFLFWQPEDEFTQIEVLSTLYRD-----NEDDLILIAIID 539
 QY 496 SYLGPS-----TSFPHRSRSPEYM-----VPLPHGGLGRSVVAYEHLAPNSYSPGHGR 545
 DB 540 RKSQKEMPKFSKSHKVELDVYKSHVPLPLP-----TLPD-----RQ 579
 QY 546 LHRQYPSLHVGRHPL-QYSPPIHGQQQLPYGI 578
 DB 580 LFQTLSPKVDTRKNILNQYVTSIFSVPFPKNV 613

RESULT 13

CING_HUMAN
 ID CING_HUMAN STANDARD; PRT; 1197 AA.
 AC Q9P2M7; Q9NR25;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Cingulin.
 GN CEN OR KIAA1319.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]_TaxID=9606;
 RP SEQUENCE FROM N.A.
 RC TISSUE=Neuroepithelium;
 RX MEDLINE=20499514; PubMed=1042084;
 RA Citi S., D'Atti F., Parry D.A.D.;
 RT "Human and Xenopus cingulin share a modular organization of the
 RT coiled-coil rod domain: predictions for intra- and intermolecular
 RT assembly.";
 RL J. Struct. Biol. 131:135-145 (2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=20181126; PubMed=10718198;
 RA Nagase T., Kikuno R., Ishikawa K.-I., Hirosewa M., Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes. XVI.
 RT The complete sequences of 150 new cDNA clones from brain which code
 RT for large proteins in vitro.";
 RL DNA Res. 7:165-73 (2000).
 RN [3]
 RP INTERACTION WITH TJPI.
 RX MEDLINE=22140336; PubMed=12023291;
 RA D'Atti F., Nadalutti F., Citi S.;
 RT "Evidence for a functional interaction between cingulin and ZO-1 in
 RT cultured cells.";
 RL J. Biol. Chem. 277:27757-27764 (2002).
 CC -!- FUNCTION: Probably plays a role in the formation and regulation of
 CC the tight junction (TJ) paracellular permeability barrier.
 CC -!- SUBUNIT: Homodimer (By similarity). Interacts with TJPI/ZO-1.
 CC -!- TISSUE SPECIFICITY: Localized on the cytoplasmic face of tight
 CC junctions of polarized epithelia and some endothelia. Expressed in
 CC pancreas, kidney, liver and lung, but not in skeletal muscle,
 CC placenta, brain or heart.
 CC -!- DOMAIN: Deletion of the ZO-1 interaction motif (ZIM) decreases but
 CC does not abolish colocalization with ZO-1.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 DR EMBL; AF263462; AAF74498.1; ALT INIT.
 DR EMBL; AB037740; BAA92557.1; ALT INIT.
 DR InterPro; IPR002928; Myosin_tail.
 DR Pfam; PF01576; Myosin_tail; 1.
 DR

KW Tight junction; Coiled coil. HEAD.
FT DOMAIN 1 351
FT DOMAIN 352 1154
FT DOMAIN 1155 1197
FT DOMAIN 363 836
FT SITE 42 56
FT SITE 106 400
FT SEQUENCE 1197 AA; 136385 MW; 0C9375283ABAAF3D CRC64;
Query Match 4.0%; Score 124.5; DB 1; Length 1197;
Best Local Similarity 19.9%; Pred. No. 5.5; Indels 143; Gaps 25;
Matches 111; Conservative 91; Mismatches 213;
QY 7 TVAAOPTTANPLLOKHOS-----EQRRR-----ELPKIV----- 36
DB 163 TIDTAPLSVDSLINKFQSGQARGTGRTRMLPEQKRKSLDRLPDTPEERE 222
QY 37 -----ETESTMDITICQSQPQLKSIDELAFSAVAVTFKQPD--LQKHISIE 88
DB 223 RQSTNHTSSTKYDNHVTGTSKQASQNSLPSGFSRS-----RQTDWVLOSFEFPRS 277
QY 89 AID-SKLENGVVLAAARNNHOPMLSPRNNSVETTV--TVSQPSQEIIVPETSNNKPEG 145
DB 278 AQDPTWLFKSPDLLRD----QEAAPGSDVDMKATIYGLREGSSE--SETSVRKV 331
QY 146 GRMCE-----LMCKGLRKIYVA-----NTSD-QAKLMEEI-----PSALKLAKEPA 186
DB 332 SLVLEKMOPLVMVSSGSTKAVAGQGLTRKVEELQRLDEEVKKRQKLEFPOVGLERQLE 391
QY 187 KPVLCIGKFKYLOGRRATTKESPMSARQVSLLLLEGFLMPDRKPKGKVKIESWIKDEAE 246
DB 392 EKTBCSLQLELLER---KGEAQSQNK-----LQNKRLLDQGED-----LRHGLE 436
QY 247 TAAVAWRKELMTTEGLAAEKMDARGLLLVACFGVPSNFRSTDLDLIRMSGNEIAGA 306
DB 437 TQWMBELQNKLVHQPPEPAKEVLLKLL-----ETRELLBEV-LECKQVVEEQ 483
QY 307 LK-RSQFLVPVTSYGIVSSIKRGWHIEALEVMYTFGMEKFSALVLTSLFKMSKESFER 365
DB 484 LRERRELTALGALKEEVSADSDQEVHVRQY-----QRDTEQLRR 525
QY 366 AKRKAQSPFAPKEAATKQAVLSSVMQCMETHKLDPAKELFQ-WQ-----IKE 412
DB 526 SMDQATQDHAVLEAERQMSALVRGLQ---RELSETSEETHQWSQFQKNKEDLRATQ 581
QY 413 QIVSLEKDTQLDQKEMEKARSLMEEAALAKRMVNOI---KRPILSPME-----MPP 464
DB 582 ELQLRMKEKEEELGKIEVLQRELEQARASAGDTRQVLEVLKELRLTQELKELQAE 641
QY 465 VTSSSYSPYTRDSRFPQ 482
DB 642 RSQEVAGRRDRDELEKQ 659
RESULT 14
PCLO CHICK STANDARD; PRT; 5120 AA.
AC 09pU36;
DT 28-FEB-2003 (rel. 41, Created)
DT 28-FEB-2003 (rel. 41, Last sequence update)
DE 28-FEB-2003 (rel. 41, Last annotation update)
DE Ficollo protein (Aczonin) (fragment).
GN PCLO OR ACZ.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=99439764; PubMed=10508862;
RA Wang X., Kibschull M., Laue M.M., Lichte B., Petrasch-Parwez E.,

RA Killmann M.W.;
RT "Aczonin, a 550-kd putative scaffolding protein of presynaptic active
RT zones, shares homology regions with rim and bassoon and binds
RT profilin";
RL J. Cell Biol. 147:151-162(1999).
CC -!- FUNCTION: May act as a scaffolding protein involved in the
CC organization of synaptic active zones and in synaptic vesicle
CC trafficking (By similarity).
CC -!- SUBUNIT: Interacts with Rabac1/Prai and profilin (By similarity).
CC -!- SUBCELLULAR LOCATION: Concentrated at the presynaptic side of
CC synaptic junctions (By similarity).
CC -!- DOMAIN: C2 domain 1 is involved in binding calcium and
CC phospholipids. Calcium binds with low affinity but with high
CC specificity and induces a large conformational change.
CC -!- SIMILARITY: Contains 2 C2 domains.
CC -!- SIMILARITY: Contains 1 PDZ/DHR domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DB EMBL; Y19187; CAB60725.1; --
DR HSSP; P04410; IAZ5.
DR GO; GO:0045202; C:synaptic junction; ISS.
DR GO; GO:0005509; F:calcium ion binding activity; ISS.
DR GO; GO:0005544; F:calcium-dependent phospholipid binding acti...; ISS.
DR GO; GO:0005522; F:profilin binding activity; ISS.
DR GO; GO:0007010; P:cytoskeleton organization and biogenesis; ISS.
DR GO; GO:0016080; P:synaptic vesicle targeting; ISS.
DR InterPro; IPR000008; C2.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001565; Synaptotagmin.
DR Pfam; PF00168; C2; 2.
DR Pfam; PF00595; PDZ; 1.
DR PRINTS; PR00399; SYNAPTOTAGMN.
DR SMART; SM00239; C2; 2.
DR SMART; SM00228; PDZ; 1.
DR PROSITE; PS00499; C2_DOMAIN_1; 1.
DR PROSITE; PS00004; C2_DOMAIN_2; 2.
DR PROSITE; PS0106; PDZ; 1.
KW Calcium/phospholipid-binding; Zinc; Metal-binding; Zinc-finger;
KW Repeat.
FT NON_TER 1 357
FT DOMAIN 258 357
FT 10 X 10 AA TANDEM APPROXIMATE REPEATS OF
FT P-A-K-P-Q-P-Q-P-X.
FT C4-TYPE (POTENTIAL).
FT C4-TYPE (POTENTIAL).
FT POLY-PRO.
FT PDZ.
FT C2 DOMAIN 1.
FT C2 DOMAIN 2.
FT SEQUENCE 5120 AA; 560751 MW; A638D9891B658412 CRC64;
Query Match 4.0%; Score 123.5; DB 1; Length 5120;
Best Local Similarity 17.9%; Pred. No. 45;
Matches 129; Conservative 115; Mismatches 296; Indels 181; Gaps 27;
QY 4 YPPTVAAOP-TTANPLLOKHOSQRRREL---PKIVETESTMDITICQSK----- 51
DB 3229 YDPSGTGSPQTITDQALLEGQYATENGQFWPTDDATTATAGVIGIEISQSQTYWTVQSD 3288
QY 52 -----QPQFLKSIDELAAFSVAVTFKR-----QFDDLQKHIES-----IEN 88
DB 3289 GITQYIPRSGILSVSEMSLKDIDVREKQKKRSMFKLRGPEELESLEEPKCYK 3348
QY 89 AIDSKLKESGVVLAAARNNHOPMLSPRNNSVETTVVSQSQ---EIVPETSNNKPEG 145
DB 3349 IVDGSGVQTDDEGDGRGYTNRR-----RRTKKSVDYTSVQTDDEQDEWDLSSRGRPRV 3403

Genome Res. 11:1053-1070(2001).

[3] SEQUENCE FROM N.A. (ISOFORM 3).

RC TISSUE=Phleochromocytoma;

RA Li Y., Peng Y., Li N., Gu W., Han Z., Fu G., Chen Z.;

RT "A novel gene expressed in human pheochromocytoma.;"

RA Submitted (SFP-1999) to the EMBL/GenBank/DBS databases.

[4] SEQUENCE FROM N.A.

RP MEDLINE=20057165; PubMed=10591208;

RA Dunham I., Hunt A.R., Collins J.E., Bruskewich R., Beare D.M.,

RA Clamp M., Smith L.J., Ainscough R., Almeida J.P., Babbage A.,

RA Baguley C., Bailey J., Barlow K., Bates K.N., Beasley O., Bird C.P.,

RA Blakey S., Bridgeman A.M., Buck D., Burgess J., Burrill W.D.,

RA Burton J., Carder C., Carter N.P., Chen Y., Clark G., Clegg S.M.,

RA Cobley G.J., Cole C.G., Collier R.E., Connor R.E., Conroy D., Corby N.,

RA Coville V.J., Cox A.V., Davis J., Dawson E., Dhami P.D., Dockree C.,

RA Dodsworth S.J., Durbin R.M., Ellington A., Evans K.L., Fey J.M.,

RA Fleming K., French L., Garner A.A., Gilbert J.G.R., Goward M.E.,

RA Graham D., Griffiths M.N., Hall C., Hall R., Hall-Tamlyn G.,

RA Heathcott R.W., Ho S., Holmes S., Hunt S.E., Jones M.C., Kershaw J.,

RA Kimberley A., King A., Laird G.K., Langford C.P., Leversha M.A.,

RA Lloyd C., Lloyd D.M., Martyn I.D., Mashreght-Mohammadi M.,

RA Matthews L., McCann O.T., McCloy J., McLaren S., McMurray A.A.,

RA Milne S.A., Mortimore B.J., Odell C.N., Pavitt R., Pearce A.V.,

RA Pearson D., Phillimore B.J., Phillips S.H., Plumb R.W., Ramsay H.,

RA Ramsey Y., Rogers L., Ross M.T., Scott C.E., Sehra H.K., Skuce C.D.,

RA Smalley S., Smith M.L., Soderlund C., Spragon L., Steward C.A.,

RA Sulston J.E., Swann R.M., Vaudin M., Wall M., Wallis J.M.,

RA Whiteley M.N., Willey D., Williams L., Williams S., Williamson H.,

RA Wilmer T.E., Wilming L., Wright C.L., Hubbard T., Bentley D.R.,

RA Beck S., Rogers J., Shimizu N., Minoshima S., Kawasaki K., Sasaki T.,

RA Asakawa S., Kudoh J., Shintani A., Shibuya K., Yoshizaki Y., Aoki N.,

RA Mitsuyma S., Roe B.A., Chen F., Chu L., Crabtree J., Deschamps S.,

RA Do A., Do T., Dorman A., Fang F., Fu Y., Hu P., Hua A., Kenton S.,

RA Lai H., Lao H.I., Lewis J., Lewis S., Lin S.-P., Loh P., Malaj E.,

RA Nguyen T., Pan H., Phan S., Qi S., Qian Y., Ray L., Ren Q., Shaull S.,

RA Sloan D., Song L., Wang Q., Wang Y., Wang Z., White J., Williamson D.,

RA Wu H., Yao Z., Zhan M., Zhang G., Chisoe S., Murray J., Miller N.,

RA Minx P., Fulton R., Johnson D., Bemis G., Bentley D., Bradshaw H.,

RA Bourne S., Cordes M., Du Z., Fulton L., Goela D., Graves T.,

RA Hawkins J., Hinds K., Kemp K., Latreille P., Layman D., Ozersky P.,

RA Rohlfing T., Scheet P., Walker C., Wansley A., Wohldmann P., Pepin K.,

RA Nelson J., Korf I., Bedell J.A., Haller L., Mardis E., Waterston R.,

RA Wilson R., Emanuel B.S., Shaikh T., Kurahashi H., Saitta S.,

RA Badar M.L., McDermid H.E., Johnson A., Wong A.C.C., Morrow B.E.,

RA Edelmann L., Kim U.J., Shiruya H., Simon M.I., Dumanski J.P.,

RA Feyraud M., Kedra D., Seroussi E., Fransson I., Tapia I., Bruder C.E.,

RA O'Brien K.P., Wilkinson P., Bodenteich A., Hartman K., Hu X.,

RA Khan A.S., Lane L., Tilahun Y., Wright H.;

RT "The DNA sequence of human chromosome 22.;"

RT Nature 402:489-495(1999).

RA Hawkins J., Hinds K., Kemp K., Latreille P., Layman D., Ozersky P.,
 RA Rohlfing T., Schect P., Walker C., Wansley A., Wohldmann P., Pepin K.,
 RA Nelson J., Kori I., Bedell J.A., Haller L., Mardis E., Waterson R.,
 RA Wilson R., Emanuel B.S., Shaikh T., Kurahashi H., Saifta S.,
 RA Budarf M.L., McDermid H.E., Johnson A., Wong A.C.C., Morrow B.E.,
 RA Edelmann L., Kim U.J., Shizuya H., Simon M.I., Dumanski J.P.,
 RA Peyraud M., Kedra D., Seroussi E., Fransson I., Tapia I., Bruder C.E.,
 RA O'Brien K.P., Wilkinson P., Bodenteich A., Hartman K., Hu X.,
 RA Khan A.S., Lane L., Tiliahun Y., Wright H.,
 RT "The DNA sequence of human chromosome 22."
 RN Nature 402:489-495(1999).
 RL [5]
 RN SEQUENCE FROM N.A. (ISOFORM 2).
 RP
 RC TISSUE=Eye, and Skin;
 RP MEDLINE=22389257; PubMed=12477932;
 RX STRAUSBERG R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.P., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diachenko L., Marusina K., Farmer A.C., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Schetz T.B.,
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahy J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J.J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalski U., Smailus D.E.,
 RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 14, 2003, 10:34:34 ; Search time 107 seconds
(without alignments)
1468.729 Million cell updates/sec

Title: US-09-890-475-1
Perfect score: 3104
Sequence: 1 MSNYPPTVAAGTTTANPLL.....RYLGLSNQRSPRNSSLDPX 609

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL.23.*
1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phage.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*
15: sp_virus.*
16: sp_bacteriap.*
17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	931	30.0	196	Q8RV21	arabidopsis
2	927	29.9	196	Q8RUS4	arabidopsis
3	921	29.7	186	Q8RU84	arabidopsis
4	921	29.7	186	Q8RU28	arabidopsis
5	921	29.7	186	Q8RWP6	arabidopsis
6	697	22.5	138	Q8RVC0	arabidopsis
7	574	18.5	104	Q8RV51	arabidopsis
8	340	11.0	558	Q8L7D6	arabidopsis
9	296.5	9.6	532	Q8LUV4	arabidopsis
10	295	9.5	473	Q8C6S2	arabidopsis
11	286.5	9.2	862	Q84649	arabidopsis
12	282	9.1	532	Q840H8	arabidopsis
13	277	8.9	505	Q8W4S7	arabidopsis
14	262.5	8.5	1337	Q8LV73	arabidopsis
15	245	7.9	470	Q8FFL1	arabidopsis
16	234.5	7.6	507	O23345	arabidopsis

17	156.5	5.0	3117	4	Q8WY20	homo sapien
18	151.5	4.9	2594	5	Q9VM88	drosohila
19	150	4.8	1088	4	Q8IUD3	homo sapien
20	149.5	4.8	948	11	Q8CIY9	rattus norv
21	146	4.7	992	4	Q8IUD5	homo sapien
22	146	4.7	1003	4	Q9UP81	homo sapien
23	145.5	4.7	948	4	Q9UIK7	homo sapien
24	143.5	4.6	1116	4	Q8IUD2	homo sapien
25	141.5	4.6	1248	3	Q8XOV7	neurospora
26	140.5	4.5	775	4	Q96CN9	homo sapien
27	139.5	4.5	1201	5	O16844	drosohila
28	138	4.4	778	11	Q8CCR3	mus musculus
29	138	4.4	1291	17	Q8T2F2	pyrococcus
30	137.5	4.4	2074	11	Q88542	mus musculus
31	137	4.4	809	11	Q9D4H2	mus musculus
32	136.5	4.4	712	5	Q8IH04	drosohila
33	136.5	4.4	1201	5	Q9V4K0	drosohila
34	135	4.3	976	11	Q99M12	mus musculus
35	135	4.3	1120	11	Q99M11	mus musculus
36	134	4.3	841	10	O23037	arabidopsis
37	132.5	4.3	1762	10	Q94DC2	oryza sativ
38	132	4.3	1316	16	Q8D2C0	wiggleswort
39	132	4.3	5327	5	O76891	drosohila
40	131.5	4.2	993	4	Q9NWP9	homo sapien
41	131.5	4.2	1249	4	Q8HGM1	homo sapien
42	130.5	4.2	981	4	O15014	homo sapien
43	130.5	4.2	5412	5	Q9W596	drosohila
44	128	4.1	2238	11	O70365	mus musculus
45	127.5	4.1	1005	4	Q9P216	homo sapien

ALIGNMENTS

RESULT 1

Q8RV21 PRELIMINARY; PRT; 186 AA.
ID Q8RV21
AC Q8RV21; 01-JUN-2002 (TREMREL. 21, Created)
DT 01-JUN-2002 (TREMREL. 21, Last sequence update)
DT 01-JUN-2002 (TREMREL. 21, Last annotation update)
DE FRIGIDA protein (Fragment).
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucoids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCB1_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. RSCH-4, cv. GOT-32, cv. TSU-0, and cv. Ler;
RA Hagenblad J., Nordborg M.;
RT "Sequence Variation and Haplotype Structure Surrounding the Flowering
Time Locus FRI in Arabidopsis thaliana";
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY092681; AAM12134.1; -
DR EMBL; AY092683; AAM12136.1; -
DR EMBL; AY092691; AAM12144.1; -
DR EMBL; AY092693; AAM12146.1; -
FT NON_TER 1
FT NON_TER 186
SQ SEQUENCE 186 AA; 20727 MW; 4916F2C7AB32228B CRC64;

Query Match 30.0%; Score 931; DB 10; Length 186;
Best Local Similarity 99.5%; Pred. No. 1.1e-57;
Matches 185; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy	34	KIVETESTMDITIGSQKQPOFLKSIDELAAFSVAVETFKQFDLDQKHIESIENADSK	93
Db	1	KIVETESTMDITIGSQKQPOFLKSIDELAAFSVAVETFKQFDLDQKHIESIENADSK	60
Qy	94	LESNGVVLAAARNNNFHQPMLSPPRNNVSVETTVTSQPSQEIIVPETSKPGGRMCMLC	153
Db	61	LESNGVVLAAARNNNFHQPMLSPPRNNVSVETTVTSQPSQEIIVPETSKPGGRMCMLC	120

```

QY 154 SKGLRKYIYANISDQAKLMEIIPSAKLAKPAKFLVDCIGKFLQGRRAFTKESPMSSA 213
Db 121 SKGLRKYIYANISDQAKLMEIIPSAKLAKPAKFLVDCIGKFLQGRRAFTKESPMSSA 180
QY 214 RQVSL 219
Db 181 RQVSL 186

RESULT 2
QBRU84
ID QBRU84 PRELIMINARY; PRT; 186 AA.
AC QBRU84;
DT 01-JUN-2002 (TRENBLrel. 21, Created)
DT 01-JUN-2002 (TRENBLrel. 21, Last sequence update)
DE FRIGIDA protein (Fragment).
OS Arabidopsis thaliana (Mouse-ear creess).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;

[1]
RN SEQUENCE FROM N.A.
RP STRAIN=cv. MT-0, cv. Koln, cv. Col, cv. NC-6, and cv. DEM-4;
RA Hagenblad J., Nordborg M.;
RT "Sequence Variation and Haplotype Structure Surrounding the Flowering
Time Locus FRI in Arabidopsis thaliana.";
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY092678; AAM12131.1; -
DR EMBL; AY092680; AAM12133.1; -
DR EMBL; AY092688; AAM12141.1; -
DR EMBL; AY092689; AAM12142.1; -
DR EMBL; AY092690; AAM12143.1; -
FT NON_TER 1
FT NON_TER 186
SQ SEQUENCE 186 AA; 20781 MW; F809E401DEDC1CDC CRC64;

Query Match 29.7%; Score 921; DB 10; Length 186;
Best Local Similarity 98.9%; Pred. No. 5.3e-57;
Matches 184; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 34 KIVETESTSMITIGQSKQPFKLSIDELAAFSVAVETFKRQFDDLKQHIESTENAI DSK 93
Db 1 KIVETESTSMITIGQSKQPFKLSIDELAAFSVAVETFKRQFDDLKQHIESTENAI DSK 60

QY 94 LESNGVLAARNNHFQPMSPRRNVSVETTVTSQPSQEIIVPETSNGKPEGGRMCELMC 153
Db 61 LESNGVLAARNNHFQPMSPRRNVSVETTVTSQPSQEIIVPETSNGKPEGGRMCELMC 120

QY 154 SKGLRKYIYANISDQAKLMEIIPSAKLAKPAKFLVDCIGKFLQGRRAFTKESPMSSA 213
Db 121 SKGLRKYIYANISDQAKLMEIIPSAKLAKPAKFLVDCIGKFLQGRRAFTKESPMSSA 180

QY 214 RQVSL 219
Db 181 RQVSL 186

RESULT 4
QBRU84
ID QBRU84 PRELIMINARY; PRT; 186 AA.
AC QBRU84;
DT 01-JUN-2002 (TRENBLrel. 21, Created)
DT 01-JUN-2002 (TRENBLrel. 21, Last sequence update)
DE FRIGIDA protein (Fragment).
OS Arabidopsis thaliana (Mouse-ear creess).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;

[1]
RN SEQUENCE FROM N.A.
RP STRAIN=cv. Lund, cv. Kent, cv. Vimmerby, cv. PU-2-8, and cv. PU-2-3;
RA Hagenblad J., Nordborg M.;
RT "Sequence Variation and Haplotype Structure Surrounding the Flowering
Time Locus FRI in Arabidopsis thaliana.";
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY092682; AAM12135.1; -
DR EMBL; AY092685; AAM12138.1; -
DR EMBL; AY092686; AAM12139.1; -
DR EMBL; AY092687; AAM12140.1; -
DR EMBL; AY092692; AAM12145.1; -
FT NON_TER 1
FT NON_TER 186
SQ SEQUENCE 186 AA; 20688 MW; 591214F7A5E24DAA CRC64;

Query Match 29.7%; Score 921; DB 10; Length 186;
Best Local Similarity 98.9%; Pred. No. 5.3e-57;
Matches 184; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 34 KIVETESTSMITIGQSKQPFKLSIDELAAFSVAVETFKRQFDDLKQHIESTENAI DSK 93
Db 1 KIVETESTSMITIGQSKQPFKLSIDELAAFSVAVETFKRQFDDLKQHIESTENAI DSK 60

```

```
QY 94 LEANGVVLARNNNFHQPMLEPPNNVSVETTTVVSQSQEIVPETSNNKPGGRVCELMC 153
|
|
|
Db 61 LEANGVVLARNNNFHQPMLEPPNNVSVETTTVVSQSQEIVPETSNNKPGGRVCELMC 120
|
|
|
QY 154 SKGLRKYIYANISDQAKLMEIIPSAKLAKPAKFVLDICIGKIFYLQGRRAFTKESPMSSA 213
|
|
|
Db 121 SKGLRKYIYANISDQAKLMEIIPSAKLAKPAKFVLDICIGKIFYLQGRRAFTKESPMSSA 180
|
|
|
QY 214 RQVSL 219
|
|
|
Db 181 RQVSL 186
|
|
|
RESULT 5
Q8RWPF6 PRELIMINARY; PRT; 186 AA.
AC Q8RWPF6;
DT 01-JUN-2002 (T-EMBLrel. 21, Created)
DT 01-JUN-2002 (T-EMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (T-EMBLrel. 21, Last annotation update)
DE FRIGIDA protein (Fragment)
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Algastrum;
RA Hagenblad J., Nordborg M.;
RT "Sequence Variation and Haplotype Structure Surrounding the Flowering
Time Locus FRI in Arabidopsis thaliana.";
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY092684; AAM12137.1; -
DR NON_TER 1
FT NON_TER 186
FT NON_TER 186
SQ SEQUENCE 186 AA; 20813 MW; 4643EF0A7ADB7C0 CRC64;
Query Match 29.7%; Score 921; DB 10; Length 186;
Best Local Similarity 98.9%; Pred. No. 5.3e-57;
Matches 184; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 34 KIVETSTSDITIGSKQPFKLSDELAAFSVAVETFKRQFDLQKHIESIENAIISK 93
|
|
|
Db 1 KIVETSTSDITIGSKQPFKLSDELAAFSVAVETFKRQFDLQKHIESIENAIISK 60
|
|
|
QY 94 LEANGVVLARNNNFHQPMLEPPNNVSVETTTVVSQSQEIVPETSNNKPGGRVCELMC 153
|
|
|
Db 61 LEANGVVLARNNNFHQPMLEPPNNVSVETTTVVSQSQEIVPETSNNKPGGRVCELMC 120
|
|
|
QY 154 SKGLRKYIYANISDQAKLMEIIPSAKLAKPAKFVLDICIGKIFYLQGRRAFTKESPMSSA 213
|
|
|
Db 121 SKGLRKYIYANISDQAKLMEIIPSAKLAKPAKFVLDICIGKIFYLQGRRAFTKESPMSSA 180
|
|
|
QY 214 RQVSL 219
|
|
|
Db 181 RQVSL 186
|
|
|
RESULT 6
Q8RWCO PRELIMINARY; PRT; 138 AA.
AC Q8RWCO;
DT 01-JUN-2002 (T-EMBLrel. 21, Created)
DT 01-JUN-2002 (T-EMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (T-EMBLrel. 21, Last annotation update)
DE FRIGIDA protein (Fragment)
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
```

```
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Various strains;
RA Hagenblad J., Nordborg M.;
RT "Sequence Variation and Haplotype Structure Surrounding the Flowering
Time Locus FRI in Arabidopsis thaliana.";
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY092437; AAM12050.1; -
DR EMBL; AY092438; AAM12051.1; -
DR EMBL; AY092439; AAM12052.1; -
DR EMBL; AY092440; AAM12053.1; -
DR EMBL; AY092441; AAM12054.1; -
DR EMBL; AY092442; AAM12055.1; -
DR EMBL; AY092443; AAM12056.1; -
DR EMBL; AY092444; AAM12057.1; -
DR EMBL; AY092445; AAM12058.1; -
DR EMBL; AY092446; AAM12059.1; -
DR EMBL; AY092447; AAM12060.1; -
DR EMBL; AY092448; AAM12061.1; -
DR EMBL; AY092449; AAM12062.1; -
DR EMBL; AY092450; AAM12063.1; -
DR EMBL; AY092451; AAM12064.1; -
DR EMBL; AY092452; AAM12065.1; -
DR EMBL; AY092453; AAM12066.1; -
DR EMBL; AY092454; AAM12067.1; -
DR EMBL; AY092455; AAM12068.1; -
DR EMBL; AY092456; AAM12069.1; -
DR NON_TER 1
FT NON_TER 138
FT NON_TER 138
SQ SEQUENCE 138 AA; 15674 MW; 17910646CD19CEEB CRC64;
Query Match 22.5%; Score 697; DB 10; Length 138;
Best Local Similarity 100.0%; Pred. No. 1.8e-41;
Matches 138; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 370 AQSPLAFKEAATKQLAIVLSSVMQCMETHKLPKELPGWIKQIVSLKDTLQDKEME 429
|
|
|
Db 1 AQSPLAFKEAATKQLAIVLSSVMQCMETHKLPKELPGWIKQIVSLKDTLQDKEME 60
|
|
|
QY 430 EKARSLSLMEEAALAKRMVNNQIKRPLSPMEMPVTTSSSYSPYRDRSPFSQDDDDQDE 489
|
|
|
Db 61 EKARSLSLMEEAALAKRMVNNQIKRPLSPMEMPVTTSSSYSPYRDRSPFSQDDDDQDE 120
|
|
|
QY 490 ISALVSSYLGPSTSPHR 507
|
|
|
Db 121 ISALVSSYLGPSTSPHR 138
|
|
|
RESULT 7
Q8RV51 PRELIMINARY; PRT; 104 AA.
AC Q8RV51;
DT 01-JUN-2002 (T-EMBLrel. 21, Created)
DT 01-JUN-2002 (T-EMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (T-EMBLrel. 21, Last annotation update)
DE FRIGIDA protein (Fragment)
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Various strains;
RA Hagenblad J., Nordborg M.;
RT "Sequence Variation and Haplotype Structure Surrounding the Flowering
Time Locus FRI in Arabidopsis thaliana.";
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY092537; AAM12090.1; -
DR EMBL; AY092538; AAM12091.1; -
DR EMBL; AY092539; AAM12092.1; -
DR EMBL; AY092540; AAM12093.1; -
DR EMBL; AY092541; AAM12094.1; -
```

QY	23	HOSBORRELPKIVETSTSDMTIGOSKQPFUKSIDELAAFSVAVET--FKQFDDLOK	81
Db	59	HELEDOKE-----YETKTR-----KAQELLEKKA-AAVEAKEKAALERLOK	99
QY	82	HTESIENAIIDSKLESNGVLAARNNNHQPMLSP-----PRN-----NVSVETTVTS	129
Db	100	KDAAMPFINSALD-----KYN--APVSPSVGCRWPQNAVEDSSNVFAADSI	149
QY	130	QP-----SCEIVP-----FTSNKPEGGRMCELMCSKGLRKIYIYANISDOAKLMEEIP	179
Db	150	NPDGIVQDVQISPMVNGYEVKAYPOLLKLGDMDSITGLHKFVSDNRKNLASLKEEIPMAF	209
QY	180	KLAKPAKPFVLDCIGKFV-----LQGRRAFTKESPMSSARQVSLILE--SLLMP-DR	230
Db	210	RAANPASLVLDLSLEGFYPMEAPTADGK-----KQANLLGMRRITCIMLMECLISLLSGLR	265
QY	231	GKGKVKIESWIKDEATAAFAVAKRLMT-EGGLAAAEKMDARGLLILIVACFGVPSNFRST	289
Db	266	NCLAVVLSONVKHRAKTIAGCNWPNLLESMDACNGNSLEHAFQLLATFAIVADFKED	325
QY	290	DLILDIRMSGNEITAGALKRSQFLVPMVSGIVESSIKRGMEIEALEWYTFGMEDEKESAA	349
Db	326	ELUKILPMVSRRRQAELCSLGLAEKMGVIEVLVNSGKQIDAVNLAFAFELTEQFSPV	385
QY	350	LVITSFLKMSKESPERAKRKAQSPLEAPKEAATKQLAVLSSVMQCMETHKLDPAKELPGWQ	409
Db	386	SLLKSYLIEARRSPQPRGNASPAVQDEFNERELIGLTKVIKICBEHSLE--EQYP---	440
QY	410	IKEQIVLSLEKUTLOLDKEMEESKARSLSMEEALAKMYNQOIKRPLSPMEHPPTSSS	469
Db	441	-----VEPLHKKRILQLEAKADKKGATEPMK-----POKPRPGA--QPRVTDNN	483
QY	470	-----YSPIYVRDSFPQRDDODEISALVSSYLGPDSTSFPHRSRRSPMYVPLPH	520
Db	484	NNNNNTKGYGRVIPER-YQYVYVDNRPFLSGPMAQPPPPPPQTYTFNP---AP-AH	538
QY	521	GGLGRSVYAYEHLAPNSY	538
Db	539	GNFYANCYQYQAPPPPPY	556

RESULT 9

Q9LUV4

ID	Q9LUV4	PRELIMINARY;	PRT;	532 AA.
AC	Q9LUV4;			
DT	01-OCT-2000 (TrEMBLrel. 15, Created)			
DT	01-OCT-2000 (TrEMBLrel. 15, Last sequence update)			
DE	01-OCT-2002 (TrEMBLrel. 22, Last annotation update)			
DE	Putative hydroxyproline-rich glycoprotein.			
GN	MCBI7.20 OR AT3G22440.			
OS	Arabidopsis thaliana (Mouse-ear cress).			
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;			
OC	euroside II; Brassicales; Brassicaceae; Arabidopsi.			
OX	NCBI_TaxID=3702;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=Columbia;			
RC	Sato S., Nakamura Y., Kaneko T., Kato T., Asamizu E., Tabata S.;			
RL	Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.			
RL	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=Columbia;			
RC	MEDLINE=20277480; PubMed=10819329;			
RX				
RT	Nakamura Y.;			
RT	"Structural analysis of Arabidopsis thaliana chromosome 3. I. Sequence			
RT	features of the regions of 4,504,864 bp covered by sixty pl and TAC			
RT	clones.";			
RL	DNA Res. 7:131-135 (2000).			
RL	[3]			
RP	SEQUENCE FROM N.A.			
RP	Yamada K., Liu S.X., Pham P.K., Banh J., Dale J.M., Goldsmith A.D.,			
RA	Jiang P.X., Lee J.M., Onodera C.S. Quach H.L., Tang C., Toriumi M.,			

SEQUENCE FROM N.A.
Yamada K., Liu S.X., Pham P.K., Banh J., Dale J.M., Goldsmith A.D.,
Jiang P.X., Lee J.M., Onodera C.S., Quach H.L., Tang C., Toriumi M.,
DNA Res. 7:131-135(2000).
[3]

[illegible]

DB 205 FLHLVAFAELGSLFETEBICDYFIILSKYQATITCKKIGLDRNRIGVLVQKFLDTGRLL 264
QY 332 EALEMVTYFGHEDKESALVLTSLFKWSKESFERAKKAQSPAFK-EAATKOLAVLSSV 390
DB 265 VAIRIYENWGEFEPVSLIKLSKNSREAKVCAGNYSILKVQNEATDKLSALRAV 324
QY 391 MQCMETHKLDPAKELPGWQIQEIVLSLEKTLQDQKEMEBKARSLSLMEBAALAKWYQ 450
DB 325 IKVVEKNIES-----EFMEE-KLEECVKELEDQAKQKKA-TKFNSPANFPQPOEQ 374
QY 451 QI--KRPRL---SPME---MPEVTSSSYSPLYDRSPSQDQDDQDEISALVSSYLGPS 501
DB 375 KYDNKRPVANGSMYENLTTPPLRPOQPPU-----LTPSQILQNPVGLSSIL-PG 428
QY 502 TSFPHRSRSPYVPLPHGGLGRSVAYEHLAPNSYSPGHGHLRQYSPSLVHGQRP 561
DB 429 VAVPYGNFRALFGSVAP--ASRPVFVQ-----QTGVG-----MPP 463
QY 562 LQYSPPIHQ 571
DB 464 PQYRPPYPQ 473

RESULT 11
ID 004649 PRELIMINARY; PRT; 862 AA.
AC 004649;
DT 01-JUL-1997 (TremBLrel. 04, Created)
DT 01-JUL-1997 (TremBLrel. 04, Last sequence update)
DT 01-DEC-2001 (TremBLrel. 19, Last annotation update)
DE A.TW021B04.8 protein.
GN A.TW021B04.8
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosid II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
[1]
RN
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Dante M., Wamsley P., Gibson A.;
RT "The sequence of A. thaliana TW021B04.";
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
[2]
RN
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Wash U;
RT "The A. thaliana Genome Sequencing Project.";
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
[3]
RN
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Waterston R.;
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF007271; AB561078.1; --
SQ SEQUENCE 862 AA; 96960 MW; A85368FF0C6D4F31 CRC64;

Query Match 9.2%; Score 286.5; DB 10; Length 862;
Best Local Similarity 22.1%; Pred. No. 1.7e-11;
Matches 132; Conservative 119; Mismatches 231; Indels 115; Gaps 21;

QY 18 PLLRQSQRRRELKPIVETESTSDITIGOSQPOFLKSIDELAAFSVAVEFKQFD 77
DB 355 FYEQKHVIR-----PRLTEPTAPSONLTVCFNMKHELTSMNLSIF-VFGETLRHKF- 407
QY 78 DLQKHIESIENAIKSLBNGVYLAARNN-----FHQPMSPRRNVSVETTVTSQ 130
DB 408 -----NQNIFFLLFKILKQPEV-----VCVPTGKQVKE 434
QY 131 PSQEIYVETSKPEGGMELMCKSG-----LRKIYVANSQAKMEEFSAKLKAKE 184
DB 435 SGADHQPDYIATHPSGTETKILNLSGSIKADMLRELVE---KQPLKESEDLSNALKCTPD 491

QY 185 PAKFVLDCIKFYQORRAFTKESPMSSARQVSLILLESFILMPDRGKGVKIESWIKDE 244
DB 492 PAKFLDTSMALCPTITEGVEFKMLITISASCSLLNLQKLLP-----KIGHPVKGD 544
QY 245 AETAAVARERLATEGLAAAEKMDARGLLLVACFGVPSNFRSTDLIDLIRMSGNEIA 304
DB 545 AKUAVTWKDI-----AKSKDQLEVICFLQFLGIFGVSEFKADLLGLLNSYQWTVS 600
QY 305 GALKRSOFLVPMVSGIVSESIKGMHIEALEMVTYFGMEDKFS-AALVLTSLFKWSKESF 363
DB 601 PDLCOFLGLDAPGFTONLIKTHRIKAIKDIYISFGMVHRFPQVSAIINDSLRITKESA 660
QY 364 ERARKAQSPFLAKFAA-TKQAVLSSVMQCHETHKLDPAKELPGWQIQEIVLSLEKDTL 422
DB 661 EKSRYEAKNSETTQVAIDRQVRLAAIKLCSCHLSESEFQJ--GDLEEQIKSLK--L 716
QY 423 QLDKEMEBKARSLSLMEBAALAKRMYNQIKRPLSPMEMPPVTSSSYPIYDRSFPQ 482
DB 717 RRNTSNGSGSGSASSKPDSTIKQ---SQTAKPPTVA-EVAPVTS-----NIPLE 761
QY 483 RDDQDEISALVSSYLGPSTSPHRSRSPYVPLPHGGLGRSV---YAYEHLA--PN 536
DB 762 PSTEAASSA-----SKPFSKQNR-----GKRSMSGNQQSGSHASHTS 802
QY 537 SYSPGHGHLRHOYSPSLVHGQRHPLQYSPPIHQOQ--LPYGIQRYVHRSPSEERY 591
DB 803 NHYPSHDYSLNQLTWPVDYDRGFTGFPNPDYNNQWQGPQGPQFYHLVQLDPRY 859

RESULT 12
ID 0940H8 PRELIMINARY; PRT; 532 AA.
AC 0940H8;
DT 01-DEC-2001 (TremBLrel. 19, Created)
DT 01-DEC-2001 (TremBLrel. 19, Last sequence update)
DT 01-OCT-2002 (TremBLrel. 22, Last annotation update)
DE Hydroxyproline-rich glycoprotein-like protein.
GN Z97337.18
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosid II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
[1]
RN
RP SEQUENCE FROM N.A.
RA Southwick A., Karlin-Neumann G., Nguyen M., Lam B., Miranda M.,
RA Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,
RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,
RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,
RA Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinozaki K.,
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
[2]
RN
RP SEQUENCE FROM N.A.
RA Southwick A., Karlin-Neumann G., Nguyen M., Lam B., Miranda M.,
RA Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,
RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,
RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,
RA Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinozaki K.,
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
[3]
RN
RP SEQUENCE FROM N.A.
RA Haas B.O., Volfovsky N., Town C.D., Troukhan M., Alexandrov N.,
RA Feldmann K.A., Flavell K.B., White O., Salzberg S.L.;
RT "Full-length messenger RNA sequences greatly improve genome
RT annotation.";
RL Genome Biol. 0:0-0(2002).
[4]
RN
RP SEQUENCE FROM N.A.
RA Brover V., Troukhan M., Alexandrov N., Lu Y.-P., Flavell R.,
RA Feldmann K.

RT "Full-length cDNA from Arabidopsis thaliana."
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AY054518; AAK96809.1; -
 DR EMBL: AY081543; AN10105.1; -
 DR EMBL: AY086472; AAM63474.1; -
 SQ SEQUENCE 532 AA; 59364 MW; C7257E0ACICE9B40 CRC64;

Query Match 9.1%; Score 282; DB 10; Length 532;
 Best Local Similarity 20.8%; Pred. No. 1.7e-11;
 Matches 126; Conservative 98; Mismatches 224; Indels 158; Gaps 21;
 QY 49 QSQOPLKSIDELAFSAVETFRQ-----FDDLOKHIESIENAIKLSNG 98
 DB :
 12 KSSQPSFFE-----FQASLMTSCNLLMKELSEHFTSMEONLMKSEALR 57
 QY 99 VVLAARNNNPHQPLGPPNNVSVETTVVS-----QSQEIVPETSINKP 143
 DB :
 58 QMTETLDNQTQSSIELLKREVTIDHSVIAAGKVEERVRAALDSEKARCGEDTGEV 117
 QY 144 EGG-----RMCELMCSKGLRKVIYANISQAKLMEIIPALAKLAKPAKFLDCIGK 195
 DB :
 118 DDGDGLLSALKSLCLKMDARGFGFVIARKKELENLRSQIPVALVDCVDPKLVLEAVSE 177
 QY 196 FYLQGRAPTKESPMSSARQVSLILLESF---LLMPDRGKGVKTESWKDEAETAFAVAV 252
 DB :
 178 VFPVDRG-GEKVSNDFGWACVVLLESILPVMVDPWVGKSLVLTVPVSKKAKELAEIW 236
 QY 253 RKELMTGGAAAEKMDARGLLLVACFGVPSNFRSTDLDLIRMSGNEIAGALKRSQF 312
 DB :
 237 KASLEERGGINVKTDPVHTFLOHLVTFGIVKK---DDLALYR---KLTVGSARWQOM 288
 QY 313 -----LVPMTSGIVESSIKRGMHIEALEMVTFTGMEKSAALVLTSLFKMSK--- 360
 DB :
 289 PKLAVSGLGQDQPMIEELIRGOQLDAVHFTFVGLVHLFPVPLKAYLRDAKATA 348
 QY 361 ---ESPERAKRKAQSPLAFKEATKQAVLSSVMQCMETHKLDPAKELPGWQIKQIVSL 417
 DB :
 349 LITDDSNNSGRSAHL-----VARKEQSALRAVLKCIIEYKLE--EEFPPENLKKRLDQL 400
 QY 418 EKTDLQDKEMEKARSLSLMEEAALAKRYNQOIKRPLS-PMEMPPVTSSSYPIYRD 476
 DB :
 401 EKTTEKRPKPAVIPAN-----KRTASVGGMPPPAKAGRIT----- 436
 QY 477 RSPFSQDDQDDEISALVSLGPGSTSPFHRSSRSPYVMPVPHGGLGRSVYAEHLAPN 536
 DB :
 437 -----NAVSSFPPTPTFIRSQSHSPQYGV-----AYTTSPT 471
 QY 537 SY---SPGHGRLHROQVSPSLVHG--ORHPLOY-----SPPI-HGQQQLPY 576
 DB :
 472 IYNSRSPY-----QYSPFAVHGSYQTSFVSPTAYGTYSFVAAPPPVPHPHHH 525
 QY 577 GIORVY 582
 DB 526 HIQAY 531

RESULT 13
 Q8W4S7 PRELIMINARY; PRT; 505 AA.
 ID Q8W4S7
 AC Q8W4S7
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE ATG14900/d13490c.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 ON NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Cheuk R., Chen J., Kim C.J., Koesema E., Meyers M.C., Banh J.,
 RA Bowser L., Carninci P., Dale J.M., Goldsmith A.D., Hayashizaki Y.,

RA Ishida J., Jiang P.X., Jones T., Kamiya A., Karlin-Neumann G.,
 RA Kawai J., Lam B., Lee J.M., Lin J., Liu S.X., Miranda M., Narusaka M.,
 RA Nguyen M., Onodera C.S., Palm C.J., Pham P.K., Quach H.L., Sakurai T.,
 RA Satou M., Seki M., Southwick A., Tang C.C., Toriumi M., Yamada K.,
 RA Yamamura Y., Yu G., Yu S., Shinoraki K., Davis R.W., Theologis A.,
 RA Ecker J.R.;
 RT "Arabidopsis cDNA clones."
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AY060551; AAL31182.1; -
 SQ SEQUENCE 505 AA; 56366 MW; 2C2DBA5DA1A23DC6 CRC64;

Query Match 8.9%; Score 277; DB 10; Length 505;
 Best Local Similarity 21.6%; Pred. No. 3.6e-11;
 Matches 122; Conservative 94; Mismatches 226; Indels 124; Gaps 20;
 QY 65 FSVAVETFRQDDLOKHIESIENAIKLSNGVLAARNNF-HQPMLSPPNNVSVE 123
 DB :
 15 FTSMEQNLKMEALRQMIETLDNQTQSSIE---LLKHREVTIDHSVEIAEGKVEERV 70
 QY 124 TTVTVSQSQEIVPETSINKPEG-----RMCELMCSKGLRKVIYANISQAKLMEI 175
 DB :
 71 AALDSLEKARCGEDTGEVDGGLLSALKSLCLKMDARGFGFVIARKKELENLRSQI 130
 QY 176 PSALKLAKPAKFLDCIGKFLQGRAPTKESPMSSARQVSLILLESF---LLMPDRGK 232
 DB :
 131 PVALVDCVDPKLVLEAVSEVFPVDRG-GEKVSNDFGWACVVLLESILPVMVDPWVGK 189
 QY 233 GKVKIESWIKDEAETAFAVAVKRLMTGEGLAEEKMDARGLLLVACFGVPSNFRSTDL 292
 DB :
 190 SRLLVTPSVKAKETAEATKASLEERGGINVKTDPVHTFLOHLVTFGIVKK---DDL 245
 QY 293 DLIRMSGNEIAGALKRSQF-----LVPMTSGIVESSIKRGMHIEALEMVTFTGME 344
 DB :
 246 ALYR---KLTVGSARWQOMPKLAVSGLGQDQPMIEELIRGOQLDAVHFTFVGLVH 301
 QY 345 KPSAALVLTSLFKMSK-----ESFERAKRKAQSPLAFKEATKQAVLSSVMQCMETH 397
 DB :
 302 LFPPVPLLKAYLRDAKATATLITDDSNNSGRSAHL-----VARKEQSALRAVLKCIIEY 355
 QY 398 KLDPAKELPGWQIKQIVSLEKDTQLDKEMEKARSLSLMEEAALAKRYNQOIKRPL 457
 DB :
 356 KLE--EEFPPENLKKRLDQLEKTKERKPAVIPAN-----KRTA 394
 QY 458 S-PMEMPPVTSSSYPIYDRSPFSQDDQDEISALVSLGPGSTSPFHRSSRSPYVMPV 516
 DB :
 395 SYSGMPPPAKAGRIT-----NAVSSFPPTPTFIRSQSHSPQYGV 435
 QY 517 PLPHGGLGRSVYAEHLAPNSY---SPGHGRLHROQVSPSLVHG--ORHPLOY----- 564
 DB :
 436 P-----AYTTSPTIYNSRSPY-----QYSPFAVHGSYQTSFVSPTAYGTY 478
 QY 565 -----SPPI-HGQQQLPYGIQVY 582
 DB 479 CSPVAAPPPVPHPHHHHIQAY 504

RESULT 14
 Q9LV73 PRELIMINARY; PRT; 1337 AA.
 ID Q9LV73
 AC Q9LV73
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE Genomic DNA, chromosome 5, Pl clone: M0E7.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 ON NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA STRAIN=Columbia;
 RA MEDLINE=20181125; PubMed=10718197;

P1 clones."
RL DNA Res. 4:215-230 (1997).
[2]
RP SEQUENCE FROM N.A.
RR Shinn P., Chen H., Cheuk R., Kim C.J., Shinn P., Banh J., Meyers M.C., Banh J.,
RA Bower L., Carinici P., Chang E., Dale J.M., Goldsmith A.D., Goldsmith A.D.,
RA Hayashizaki Y., Ishida J., Jones T., Kamiya A., Karlin-Neumann G., Karlin-Neumann G.,
RA Kawai J., Lam B., Lee J.M., Lin J., Miranda M., Narusaka M., Narusaka M.,
RA Nguyen M., Onodera C.S., Palm C.J., Quach H.L., Sakurai T., Satou M.,
RA Seki M., Southwick A., Tang C.C., Toriumi M., Wu H.C., Yamada K.,
RA Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,
RA Ecker J.R.;
RA "Arabidopsis cDNA clones."
RRT Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
RRL [3]
RRP SEQUENCE FROM N.A.
RRR Cheuk R., Chen H., Kim C.J., Shinn P., Banh J., Bowser L.,
RA Carnici P., Chang E., Dale J.M., Goldsmith A.D., Hayashizaki Y.,
RA Ishida J., Jones T., Kamiya A., Karlin-Neumann G., Kawai J., Lam B.,
RA Lee J.M., Lin J., Miranda M., Narusaka M., Nguyen M., Onodera C.S.,
RA Palm C.J., Quach H.L., Sakurai T., Satou M., Seki M., Southwick A.,
RA Tang C.C., Toriumi M., Wu H.C., Yamada K., Yamamura Y., Yu G., Yu S.,
RA Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;
RRT "Arabidopsis ORF clones."
RRL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
RRP EMBL; AB005242; BAB09599.1; -
RRR EMBL; AY075663; AAL77670.1; -
RRL EMBL; AY101525; AAM26646.1; -
RRP SEQUENCE 470 AA; 52831 MW; 2924997BB01366D6 CRC64;
Query Match 7.9%; Score 246; DB 10; Length 470;
Best: Local Similarity 22.5%; Pred. No. 4.9e-09;
Matches 125; Conservative 92; Mismatches 204; Indels 134; Gaps 22;
QY 56 LKSIDELAAPSVAVETFKQFDLQKHTESE-----NAIDSKLSNGVLAARNNF 108
Ddb 11 INQIDE-----KKEKKAFDFDQAHSLSPSPSLSWSSEIDSHFSLOSSLASREPLL 64
QY 109 HQPMLSPRRNVSVETVT--VSOPSOEIVPETSINKPEGGMCELMCSGLRKYIYANI 165
Ddb 65 HS---TSPLEHDSYRIDASDAGKSSSESEVBPVPELRALCEKDIGLIKYLIRIW 121
QY 166 SDQAKLMEETPSALKAKEPAKFLVDCI-GKFLQGRAPFTKESPMGS-----ARQVSLL 219
Ddb 122 DDETPLNQVSAAIRVSPDTASWLDLEGSY-----TPSSGSGSFVRRRVFVL 171
QY 220 ILSEFLMPDRGRGKYKIESWKDEAETAANAWAREKRLMTEGGLAAAEKMDARGLLLVAC 279
Ddb 172 LMEVLTEI---NANITVDT--RNRAKKLYAHWKSIV-----GVKPFPEALVFLHVA 218
QY 280 FGVPSPFRSTDLLLRMSG-----NETAGALKBSQLFVPMVSGIVESSIKRGWHI 331
Ddb 219 FELGSEFDEEUSDYVFMATKYQATLVCNKIGVDKRE-----VOKLITLDSKPI 271
QY 332 EALEMYVTQMEDKFSAAVLVLTSPKMSKESFERA-----KRKAQSPFLAPKAAATKQLA 385
Ddb 272 LAVKFMVYECGMDTEFEPPIVLVASYIKDCREALRVCVEDNYSLKSN-----BASDKV 326
QY 386 VLSSVQCWETHKLDPAKELPCQWJKEQIVSLKEDTLQDKEMEEKARSLSLMEERALAK 445
Ddb 327 ALKPLKIKDQNL--SEPTQEKVEERVELEK-----KALRK 364
QY 446 RMYNQIKRPRISPMEMPVPTSSSYPIYRDRSPFSQBDDBQDEISALVSSYLGPTSTFP 505
Ddb 365 R---NTTNPPKBPQKGGKRT-----RDCXGSGVPEPQQLSRPEALLMPEHS-- 412
QY 506 HRSRSPPEYVPLPHGGLGRSVY---AYEHLAPNSYSPGHGHLRHQYSPSLVHGQRH 560
Ddb 413 -----HGLQNLPIGLMTSAFSGVVNPLTGLFGSGATPQ---SLYYAQQT 455
QY 561 ----PLQYSPPIHQ 571
Ddb 456 GYVLPPQYHPHYYSQ 470

Search completed: August 14, 2003, 10:40:00
Job time : 110 secs

PR	21-JUL-1999;	99US-0145086.	PR	21-OCT-1999;	99US-0160741.
PR	21-JUL-1999;	99US-0145088.	PR	21-OCT-1999;	99US-0160767.
PR	22-JUL-1999;	99US-0145085.	PR	21-OCT-1999;	99US-0160768.
PR	22-JUL-1999;	99US-0145087.	PR	21-OCT-1999;	99US-0160770.
PR	22-JUL-1999;	99US-0145089.	PR	21-OCT-1999;	99US-0160814.
PR	22-JUL-1999;	99US-0145192.	PR	21-OCT-1999;	99US-0160815.
PR	23-JUL-1999;	99US-0145145.	PR	22-OCT-1999;	99US-0160980.
PR	23-JUL-1999;	99US-0145218.	PR	22-OCT-1999;	99US-0160981.
PR	23-JUL-1999;	99US-0145224.	PR	22-OCT-1999;	99US-0160989.
PR	26-JUL-1999;	99US-0145276.	PR	25-OCT-1999;	99US-0161404.
PR	27-JUL-1999;	99US-0145913.	PR	25-OCT-1999;	99US-0161405.
PR	27-JUL-1999;	99US-0145918.	PR	25-OCT-1999;	99US-0161406.
PR	27-JUL-1999;	99US-0145919.	PR	26-OCT-1999;	99US-0161359.
PR	28-JUL-1999;	99US-0145951.	PR	26-OCT-1999;	99US-0161360.
PR	02-AUG-1999;	99US-0146386.	PR	26-OCT-1999;	99US-0161361.
PR	02-AUG-1999;	99US-0146388.	PR	28-OCT-1999;	99US-0161920.
PR	02-AUG-1999;	99US-0146389.	PR	28-OCT-1999;	99US-0161992.
PR	03-AUG-1999;	99US-0147038.	PR	28-OCT-1999;	99US-0161993.
PR	04-AUG-1999;	99US-0147204.	PR	29-OCT-1999;	99US-0162142.
PR	04-AUG-1999;	99US-0147302.			
PR	05-AUG-1999;	99US-0147192.			
PR	05-AUG-1999;	99US-0147260.			
PR	06-AUG-1999;	99US-0147303.			
PR	06-AUG-1999;	99US-0147416.			
PR	09-AUG-1999;	99US-0147493.			
PR	10-AUG-1999;	99US-0147935.			
PR	11-AUG-1999;	99US-0148171.			
PR	11-AUG-1999;	99US-0148319.			
PR	12-AUG-1999;	99US-0148341.			
PR	13-AUG-1999;	99US-0148565.			
PR	13-AUG-1999;	99US-0148684.			
PR	16-AUG-1999;	99US-0149368.			
PR	17-AUG-1999;	99US-0149175.			
PR	18-AUG-1999;	99US-0149426.			
PR	20-AUG-1999;	99US-0149722.			
PR	20-AUG-1999;	99US-0149723.			
PR	20-AUG-1999;	99US-0149929.			
PR	23-AUG-1999;	99US-0149902.			
PR	23-AUG-1999;	99US-0149930.			
PR	25-AUG-1999;	99US-0150566.			
PR	26-AUG-1999;	99US-0150884.			
PR	27-AUG-1999;	99US-0151065.			
PR	27-AUG-1999;	99US-0151066.			
PR	27-AUG-1999;	99US-0151080.			
PR	30-AUG-1999;	99US-0151303.			
PR	31-AUG-1999;	99US-0151438.			
PR	01-SEP-1999;	99US-0151930.			
PR	07-SEP-1999;	99US-0152363.			
PR	10-SEP-1999;	99US-0153070.			
PR	13-SEP-1999;	99US-0153758.			
PR	15-SEP-1999;	99US-0154018.			
PR	16-SEP-1999;	99US-0154039.			
PR	20-SEP-1999;	99US-0154779.			
PR	22-SEP-1999;	99US-0155139.			
PR	23-SEP-1999;	99US-0155486.			
PR	24-SEP-1999;	99US-0155659.			
PR	28-SEP-1999;	99US-0156458.			
PR	29-SEP-1999;	99US-0156596.			
PR	04-OCT-1999;	99US-0157117.			
PR	05-OCT-1999;	99US-0157753.			
PR	06-OCT-1999;	99US-0157865.			
PR	07-OCT-1999;	99US-0158029.			
PR	12-OCT-1999;	99US-0158232.			
PR	13-OCT-1999;	99US-0158369.			
PR	13-OCT-1999;	99US-0159293.			
PR	13-OCT-1999;	99US-0159294.			
PR	13-OCT-1999;	99US-0159295.			
PR	14-OCT-1999;	99US-0159329.			
PR	14-OCT-1999;	99US-0159330.			
PR	14-OCT-1999;	99US-0159331.			
PR	14-OCT-1999;	99US-0159637.			
PR	14-OCT-1999;	99US-0159638.			
PR	18-OCT-1999;	99US-0159584.			
Query Match 10.8%; Score 336; DB 21; Length 454;			Query Match 10.8%; Score 336; DB 21; Length 454;		
Best Local Similarity 25.5%; Pred. No. 2.4e-21;			Best Local Similarity 25.5%; Pred. No. 2.4e-21;		
Matches 126; Conservative 84; Mismatches 199; Indels 86; Gaps 19;			Matches 126; Conservative 84; Mismatches 199; Indels 86; Gaps 19;		
QY	85	SIENAIKSKLESXGVVLAAARNNNFHQFMILSP-----PRN-----NVSVETTVTSOP- 131	QY	85	SIENAIKSKLESXGVVLAAARNNNFHQFMILSP-----PRN-----NVSVETTVTSOP- 131
Db	3	TINSALD-----KYNN--APVSKPSVGERWPQNAVEDSSNVFAADSIIDNDPD 48	Db	3	TINSALD-----KYNN--APVSKPSVGERWPQNAVEDSSNVFAADSIIDNDPD 48
QY	132	-----SQEIVP-----ETSNKEPEGRCMELCKGLKRYIYANISDOAKLMEIIPSAKLA 182	QY	132	-----SQEIVP-----ETSNKEPEGRCMELCKGLKRYIYANISDOAKLMEIIPSAKLA 182
Db	49	GIVODVQISPMGNYEVKAYPQLLKCGDMDSGLHKFVSDNRKNLASKEEIPWAFRAA 108	Db	49	GIVODVQISPMGNYEVKAYPQLLKCGDMDSGLHKFVSDNRKNLASKEEIPWAFRAA 108
QY	183	KEPAKFVLDICIGFY-----LQGRAFTKESPMSSARQVSLILE--SFLMP-DRGKG 233	QY	183	KEPAKFVLDICIGFY-----LQGRAFTKESPMSSARQVSLILE--SFLMP-DRGKG 233
Db	109	ANPASLVLDLSLEGFYPMAPTADGK---KDNLLGMRTICMLMECLISLLSGLDRNSL 164	Db	109	ANPASLVLDLSLEGFYPMAPTADGK---KDNLLGMRTICMLMECLISLLSGLDRNSL 164
QY	234	KVIESWIKDEAETAAVAVWKRMT-EGGLAAABKMDARGILLLVACFGVPSNFRSTDLL 292	QY	234	KVIESWIKDEAETAAVAVWKRMT-EGGLAAABKMDARGILLLVACFGVPSNFRSTDLL 292
Db	165	AVLSQNVKRAKTIAGWNPLLESMDACNGNSLEAHAFIQLLATFAIVADFKEDELL 224	Db	165	AVLSQNVKRAKTIAGWNPLLESMDACNGNSLEAHAFIQLLATFAIVADFKEDELL 224
QY	293	DLIRMGSNBIAGALKRSQFLVPMVSGIVESSIKRGMHTEALEMVTYFGMEDKFSAAVLV 352	QY	293	DLIRMGSNBIAGALKRSQFLVPMVSGIVESSIKRGMHTEALEMVTYFGMEDKFSAAVLV 352
Db	225	KLI PMVSRRRQAEELCSLGLAEKMPGVIEVLVNSKQIDAVNLAFAFELTQSPVSILL 284	Db	225	KLI PMVSRRRQAEELCSLGLAEKMPGVIEVLVNSKQIDAVNLAFAFELTQSPVSILL 284
QY	353	TSFLMSKESFERAKQAQSPFAKKAATQKQAVLSVMCMQETHKLDPAKELPWQIKE 412	QY	353	TSFLMSKESFERAKQAQSPFAKKAATQKQAVLSVMCMQETHKLDPAKELPWQIKE 412
Db	285	KSYLIEARRSSPOGRPGNASPAVDDEFNERELIGLTKVICKIEEHSLE--EQYP----- 336	Db	285	KSYLIEARRSSPOGRPGNASPAVDDEFNERELIGLTKVICKIEEHSLE--EQYP----- 336
QY	413	QIVSLEKDTLQDKEMEEKARSLSLMEEAALAKRMYNQOIKRPLSPMBMPVTSSS--- 469	QY	413	QIVSLEKDTLQDKEMEEKARSLSLMEEAALAKRMYNQOIKRPLSPMBMPVTSSS--- 469
Db	337	-VEPLHRIIQLQEKAKADKGRATEPMK-----POKPRGA---QPRVTNNNNI 382	Db	337	-VEPLHRIIQLQEKAKADKGRATEPMK-----POKPRGA---QPRVTNNNNI 382
QY	470	-----YSPIYRDRSPSQRDDQDEISALVSSYLGPSSTFPHRRRRSPPEYVPLPHGL 523	QY	470	-----YSPIYRDRSPSQRDDQDEISALVSSYLGPSSTFPHRRRRSPPEYVPLPHGL 523
Db	383	NNKTYGGRVIPER-YQYVYDNRNRPFLSGPIMAAQPPPPPPPTVTNFP---AP-AHGNF 437	Db	383	NNKTYGGRVIPER-YQYVYDNRNRPFLSGPIMAAQPPPPPPPTVTNFP---AP-AHGNF 437
QY	524	GRSVYAYEHLAPNSY 538	QY	524	GRSVYAYEHLAPNSY 538
Db	438	YANCYQCAPPPPPY 452	Db	438	YANCYQCAPPPPPY 452
RESULT 7			RESULT 7		
AAG39609			AAG39609		
ID	AAG39609 standard; Protein; 456 AA.		ID	AAG39609 standard; Protein; 456 AA.	
XX	AAG39609;		XX	AAG39609;	
AC	AAG39609;		AC	AAG39609;	
DT	18-OCT-2000 (first entry)		DT	18-OCT-2000 (first entry)	
XX	Arabidopsis thaliana protein fragment SEQ ID NO: 49034.		XX	Arabidopsis thaliana protein fragment SEQ ID NO: 49034.	
DE	Protein identification; signal transduction pathway; metabolic pathway;		DE	Protein identification; signal transduction pathway; metabolic pathway;	
XX	hybridisation assay; genetic mapping; gene expression control; promoter;		XX	hybridisation assay; genetic mapping; gene expression control; promoter;	
KW			KW		

```
Db      90  KDAAMFTINSALD-----KYN--APVSKPSVGERWPCNAVESSNVFAADSTDD 139
Qy      130  QP-----SDEIVP-----ETSKPGRGCMELMCKGLRKIYIYANISDOAKLMEEIPFSA 179
Db      140  NPDGIVQDVQISPMVNCYEVKAYPQLLKCGMDSTGLHKFVSDNRKNLASLKEEIPVAF 199
Qy      180  KLAKEPAKVLDCIGKPY-----LQRRATKESPMGSAQVSLILLE--SFLIMP-DR 230
Db      200  RAAANPASLVLSLSEGFYPMFAPTADGK-----KDALNLMGERTCIMLMCELSILLSLGLDR 255
Qy      231  GKGKVKIESWIDEAETAFAVAKELMT--EGGLAAAEKMDARGILLIVACFGVPSNFRST 289
Db      256  NCLAVVLSQNVKRAKTIAGWNPLLESILDMDACNGNSLEAHAFIQLLATPAIVADFKED 315
Qy      290  DILLIRMGSNEIAGALKRSQFLPMVSGIVSESSIKRGMIEIALEMVYTFGMEDKFSAA 349
Db      316  ELIKLIPMYRRROAELCRSLGAEKMPGVIEVNVNSGKIDAVNALAFAPELTEQSPV 375
Qy      350  LVLTSPFKMSKESFERAKRKAQPL--AFKEAATKQALAVLSVMQCMETHKLDPAKPLG 407
Db      376  SLLKSLVIEARSSPQGRPNASPAVQASDEFNEBELGLKTVIKCIEHSLE--EQVP- 432
Qy      408  WQKEQIVLEKDTLQDKEMEBEKARSLSLMEEAALAKRMYNQOIKRPRLSPMEMPPVTS 467
Db      433  -----VEPLHKRILQLEKAKADKRATEPMK-----POPKRPGA---QPRVTD 473
Qy      468  SS-----YPIYDRSPFSQDDDDDETSAVSSVSGPSTSFPHRSRRSPFYWVPL 518
Db      474  NNNNNNTGTGRVIPER-YQYVYDNRPLSGFIMAAQPPPPPPPTOTYFNP---AP- 528
Qy      519  PHGGLGRSVYAYEHLAPNSY 538
Db      529  AHGNFYANCYQYQAPPPPY 548

RESULT 6
AAG06059
ID  AAG06059 standard; Protein; 454 AA.
XX
AC  AAG06059;
XX
DT  17-OCT-2000 (first entry)
XX
DE  Arabidopsis thaliana protein fragment SEQ ID NO: 2693.
KW  Protein identification; signal transduction pathway; metabolic pathway;
KW  hybridisation assay; genetic mapping; gene expression control; promoter;
KW  termination sequence.
XX
OS  Arabidopsis thaliana.
XX
PN  EF1033405-A2.
XX
PD
XX
PF  06-SEP-2000.
XX
PF  25-FEB-2000; 2000EP-0301439.
XX
PR  25-FEB-1999; 99US-0121825.
PR  05-MAR-1999; 99US-0123180.
PR  09-MAR-1999; 99US-0123548.
PR  23-MAR-1999; 99US-0125788.
PR  25-MAR-1999; 99US-0126264.
PR  29-MAR-1999; 99US-0126785.
PR  01-APR-1999; 99US-0127462.
PR  06-APR-1999; 99US-0128234.
PR  08-APR-1999; 99US-0128714.
PR  16-APR-1999; 99US-0129845.
PR  19-APR-1999; 99US-0130077.
PR  21-APR-1999; 99US-0130449.
PR  23-APR-1999; 99US-0130510.
PR  28-APR-1999; 99US-0130891.
PR  28-APR-1999; 99US-0131449.
Pr      30-APR-1999; 99US-0132048.
Pr      30-APR-1999; 99US-0132407.
Pr      04-MAY-1999; 99US-0132484.
Pr      05-MAY-1999; 99US-0132485.
Pr      06-MAY-1999; 99US-0132486.
Pr      06-MAY-1999; 99US-0132487.
Pr      07-MAY-1999; 99US-0132863.
Pr      11-MAY-1999; 99US-0134256.
Pr      14-MAY-1999; 99US-0134218.
Pr      14-MAY-1999; 99US-0134219.
Pr      14-MAY-1999; 99US-0134221.
Pr      14-MAY-1999; 99US-0134370.
Pr      18-MAY-1999; 99US-0134768.
Pr      19-MAY-1999; 99US-0134941.
Pr      20-MAY-1999; 99US-0135124.
Pr      21-MAY-1999; 99US-0135353.
Pr      24-MAY-1999; 99US-0135629.
Pr      25-MAY-1999; 99US-0136021.
Pr      27-MAY-1999; 99US-0136392.
Pr      28-MAY-1999; 99US-0136782.
Pr      01-JUN-1999; 99US-0137222.
Pr      03-JUN-1999; 99US-0137528.
Pr      04-JUN-1999; 99US-0137502.
Pr      07-JUN-1999; 99US-0137724.
Pr      08-JUN-1999; 99US-0138094.
Pr      10-JUN-1999; 99US-0138540.
Pr      10-JUN-1999; 99US-0138847.
Pr      14-JUN-1999; 99US-0139119.
Pr      16-JUN-1999; 99US-0139452.
Pr      16-JUN-1999; 99US-0139453.
Pr      17-JUN-1999; 99US-0139492.
Pr      18-JUN-1999; 99US-0139454.
Pr      18-JUN-1999; 99US-0139455.
Pr      18-JUN-1999; 99US-0139462.
Pr      18-JUN-1999; 99US-0139463.
Pr      18-JUN-1999; 99US-0139750.
Pr      18-JUN-1999; 99US-0139763.
Pr      21-JUN-1999; 99US-0139817.
Pr      22-JUN-1999; 99US-0139899.
Pr      23-JUN-1999; 99US-0140353.
Pr      23-JUN-1999; 99US-0140354.
Pr      24-JUN-1999; 99US-0140695.
Pr      28-JUN-1999; 99US-0140823.
Pr      29-JUN-1999; 99US-0140991.
Pr      30-JUN-1999; 99US-0141287.
Pr      01-JUL-1999; 99US-0141842.
Pr      01-JUL-1999; 99US-0142154.
Pr      02-JUL-1999; 99US-0142055.
Pr      06-JUL-1999; 99US-0142390.
Pr      08-JUL-1999; 99US-0142803.
Pr      09-JUL-1999; 99US-0142920.
Pr      12-JUL-1999; 99US-0142977.
Pr      13-JUL-1999; 99US-0143542.
Pr      14-JUL-1999; 99US-0143624.
Pr      15-JUL-1999; 99US-0144005.
Pr      16-JUL-1999; 99US-0144085.
Pr      16-JUL-1999; 99US-0144086.
Pr      19-JUL-1999; 99US-0144325.
Pr      19-JUL-1999; 99US-0144331.
Pr      19-JUL-1999; 99US-0144332.
Pr      19-JUL-1999; 99US-0144333.
Pr      19-JUL-1999; 99US-0144334.
Pr      19-JUL-1999; 99US-0144335.
Pr      20-JUL-1999; 99US-0144352.
Pr      20-JUL-1999; 99US-0144632.
Pr      20-JUL-1999; 99US-0144684.
Pr      21-JUL-1999; 99US-0144814.
```

PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 22-JUN-1999; 99US-0140353.
PR 22-JUN-1999; 99US-0140354.
PR 22-JUN-1999; 99US-0140695.
PR 22-JUN-1999; 99US-0140823.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 08-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.

PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156658.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 10.9%; Score 337; DB 21; Length 550;

Best Local Similarity 25.2%; Pred. No. 2.7e-21;

Matches 141; Conservative 93; Mismatches 222; Indels 104; Gaps 23;

QY 23 HQSEORRELPKIVETESTSMDITIGSKQPFKXSIDLAAFSAVET-FKQFDDLOK 81

Db 49 HELEDQKE---YETKTR-----KQELLEKKKAARAKEKAALERLOK 89

QY 82 HIESIENAIKSLKESNGVLAARNNNHQPMLSP-----PRN-----NVSVETTUTVS 129

PR	04-OCT-1999;	99US-0157117.	
PR	05-OCT-1999;	99US-0157753.	
PR	06-OCT-1999;	99US-0157865.	
PR	07-OCT-1999;	99US-0158029.	
PR	08-OCT-1999;	99US-0158232.	
PR	12-OCT-1999;	99US-0158369.	
PR	13-OCT-1999;	99US-0159293.	
PR	13-OCT-1999;	99US-0159294.	
PR	13-OCT-1999;	99US-0159295.	
PR	14-OCT-1999;	99US-0159329.	
PR	14-OCT-1999;	99US-0159330.	
PR	14-OCT-1999;	99US-0159331.	
PR	14-OCT-1999;	99US-0159637.	
PR	14-OCT-1999;	99US-0159638.	
PR	18-OCT-1999;	99US-0159584.	
PR	21-OCT-1999;	99US-0160741.	
PR	21-OCT-1999;	99US-0160767.	
PR	21-OCT-1999;	99US-0160768.	
PR	21-OCT-1999;	99US-0160770.	
PR	21-OCT-1999;	99US-0160814.	
PR	21-OCT-1999;	99US-0160815.	
PR	22-OCT-1999;	99US-0160880.	
PR	22-OCT-1999;	99US-0160981.	
PR	22-OCT-1999;	99US-0160989.	
PR	25-OCT-1999;	99US-0161404.	
PR	25-OCT-1999;	99US-0161405.	
PR	25-OCT-1999;	99US-0161406.	
PR	26-OCT-1999;	99US-0161359.	
PR	26-OCT-1999;	99US-0161360.	
PR	26-OCT-1999;	99US-0161361.	
PR	28-OCT-1999;	99US-0161920.	
PR	28-OCT-1999;	99US-0161992.	
PR	28-OCT-1999;	99US-0161993.	
PR	29-OCT-1999;	99US-0162142.	

Query Match

Best Local Similarity

Matches 141; Conservative

10.9%; Score 338; DB 21; Length 560;

25.2%; Pred. No. 2.2e-21;

93; Mismatches 222; Indels 104; Gaps 23;

QY	23	HQSEQRRLPKIVETESMBDITIGSQKQOFLKSIDELAAPSVAVET-FKQFDDLOK 81
DB	59	HELEDOEKE-----YEKTR-----KAQELLEKKAAVEAKEKAALERLOK 99
QY	82	HIESINADSKLESNGVLAANNFHPMLSP-----PRN-----NYSVETTVTS 129
DB	100	KRDAAMFTINSALD-----KYNN--APVSKPSVGERWPQNAVESDNVFAADSIIDD 149
QY	130	QP-----SOEIVP-----BTSNKPGEGRMCELMCKGLRKYIYANISQAKLMBEIPSAI 179
DB	150	NPDGIIVQDVQISPMGNYEVKAYPQLLXLCGDMDSGLHFKFVSDNKKLASLKEEIPWAF 209
QY	180	KLAKEPAKFVLDICIGKPY-----LQGRRAFTKESPMSSARQVUSLLILE--SFLIMP-DR 230
DB	210	RAAANPASVLDSLEGFPYMEAPTADGK----KDALNLLGNRETCTIMLMCELSILLSGILDR 265
QY	231	GKGKVKIESIKDEAETAFAVANKRLMT-EGGLAAAEKMDARGLLLVACFGVPSNFRST 289
DB	266	NCLAVLVSQNVKRAKTIAGWNPLESIDMACNLSLEAHAFQLLIATPAIVADFKED 325
QY	290	DLIDLIMSGNSRIAGALKESQFLVPMVSGIVESSIKRGHIEALEMYVTFGMEDKXFAA 349
DB	326	ELLKLI PMVSRRAQAEBCRLSLGAEKMPGIVELVNSGKQIDAVNLAPAFELTQFSPV 385
QY	350	LVLTSFLKMSKEFERAKRAQSPQL--APKEATKOLAVLSSVMOCMETHKLDPAKEIPG 407
DB	386	SLLSYLIIEARRSPQPGNAPQVADSEDFNERELIGLTKVICKIEHSLE--EQTP- 442
QY	408	WQIKEQIVSLEKDTQLQDKEMEKEKARSLIMEBAALAKRMYNQIKRPLSPMEMPVPTS 467
DB	443	-----VEPLHKRLIQLEKAKADKGRATEPMK-----PQKRPGEA---QPRVTD 483
QY	468	SS-----YSPYIDRDFPQRDDQDEIGALVSSVLGFTSTFPFHSRSPSPYMWPL 518

Db	484	NNNNNNKTYGRVIPER-YPOYVVDNRPFLLSGPIAAQPPPPPPQTYTFNP----	AP- 538
QY	519	PHGGLGRSVYAYEHLAPNSY 538	
Db	539	AHGNFYANCYQYQAPPFPFY 558	
RESULT 5			
AAG39608			
ID	AAG39608 standard; Protein: 550 AA.		
XX	AAG39608;		
AC	XX		
XX	XX		
DT	18-OCT-2000 (first entry)		
XX	XX		
DE	Arabidopsis thaliana protein fragment SEQ ID NO: 49033.		
XX	XX		
KW	Protein identification; signal transduction pathway; metabolic pathway;		
KW	hybridisation assay; genetic mapping; gene expression control; promoter;		
XX	XX		
OS	Arabidopsis thaliana.		
XX	XX		
DN	EP1033405-A2.		
XX	XX		
PD	06-SEP-2000.		
XX	XX		
PF	25-FEB-2000; 2000EP-0301439.		
XX	XX		
PR	25-FEB-1999;	99US-0121825.	
PR	05-MAR-1999;	99US-0123180.	
PR	09-MAR-1999;	99US-0123548.	
PR	23-MAR-1999;	99US-0125788.	
PR	25-MAR-1999;	99US-0126264.	
PR	29-MAR-1999;	99US-0126785.	
PR	01-APR-1999;	99US-0127462.	
PR	06-APR-1999;	99US-0128234.	
PR	08-APR-1999;	99US-0128714.	
PR	16-APR-1999;	99US-0129845.	
PR	19-APR-1999;	99US-0130077.	
PR	21-APR-1999;	99US-0130449.	
PR	23-APR-1999;	99US-0130510.	
PR	23-APR-1999;	99US-0130891.	
PR	30-APR-1999;	99US-0131449.	
PR	30-APR-1999;	99US-0132048.	
PR	04-MAY-1999;	99US-0132407.	
PR	05-MAY-1999;	99US-0132484.	
PR	06-MAY-1999;	99US-0132485.	
PR	06-MAY-1999;	99US-0132486.	
PR	07-MAY-1999;	99US-0132863.	
PR	11-MAY-1999;	99US-0132866.	
PR	14-MAY-1999;	99US-0134218.	
PR	14-MAY-1999;	99US-0134219.	
PR	14-MAY-1999;	99US-0134221.	
PR	14-MAY-1999;	99US-0134370.	
PR	18-MAY-1999;	99US-0134768.	
PR	19-MAY-1999;	99US-0134941.	
PR	20-MAY-1999;	99US-0135124.	
PR	21-MAY-1999;	99US-0135353.	
PR	24-MAY-1999;	99US-0135629.	
PR	25-MAY-1999;	99US-0136021.	
PR	27-MAY-1999;	99US-0136392.	
PR	28-MAY-1999;	99US-0136782.	
PR	01-JUN-1999;	99US-0137222.	
PR	03-JUN-1999;	99US-0137528.	
PR	04-JUN-1999;	99US-0137502.	
PR	07-JUN-1999;	99US-0137724.	
PR	08-JUN-1999;	99US-0138094.	
PR	10-JUN-1999;	99US-0138540.	
PR	10-JUN-1999;	99US-0138847.	
PR	14-JUN-1999;	99US-0139119.	
PR	16-JUN-1999;	99US-0139452.	

PR 25-FEB-1999; 99US-011825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 21-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140033.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142801.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148564.
PR 16-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151066.
PR 30-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.

the flowering time of a plant -
 Claim 30; Fig 6; 73pp; English.
 The present sequence represents a polypeptide capable of specifically altering the flowering time of a plant. The polypeptide is encoded by the FRI (one locus-FRIGIDA) locus of Arabidopsis. The FRI polynucleotide is used to transform plants, so that the flowering time of a plant is altered. This is used, for example, for plants in which the leaves or tubers are a commercial product, where it is desirable to avoid 'blotting' (initiation of flowers and stem elongation) at too early a stage. Conversely, it may be desirable to alter flowering under certain circumstances e.g. to vary flower production across the seasons.

XX Sequence 609 AA;
 Query Match 100.0%; Score 3104; DB 21; Length 609;
 Best Local Similarity 100.0%; Pred. No. 1.6e-272;
 Matches 609; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSNYPPTVAQAQPTTANPLLRHQSEORRELPKIVETESTSMDIITIGOSKQFQFKSID 60
 DB 1 MSNYPPTVAQAQPTTANPLLRHQSEORRELPKIVETESTSMDIITIGOSKQFQFKSID 60
 QY 61 ELAASFVAVETFKQPDLDLOKHIESIENADSKLENGVLAARNNPHQMLSPRNNV 120
 DB 61 ELAASFVAVETFKQPDLDLOKHIESIENADSKLENGVLAARNNPHQMLSPRNNV 120
 QY 121 SVETTVTVSQSQIIVPETNKGGRMCMLCKGLRKVIYANISDQAKLMEETPSALK 180
 DB 121 SVETTVTVSQSQIIVPETNKGGRMCMLCKGLRKVIYANISDQAKLMEETPSALK 180
 QY 181 LAKSPAKFVLDICIKFYLOQRRAFTKESPMSSARQVSLILFSLMPDRGKGVKIESW 240
 DB 181 LAKSPAKFVLDICIKFYLOQRRAFTKESPMSSARQVSLILFSLMPDRGKGVKIESW 240
 QY 241 IKDRAETAAVAWRKRLMTEGGLAAAEKMDARGLLLVACFGVSNFRSTDLLDLIRMSG 300
 DB 241 IKDRAETAAVAWRKRLMTEGGLAAAEKMDARGLLLVACFGVSNFRSTDLLDLIRMSG 300
 QY 301 NEIAGALKRSQFLVPMVSGIVESSIKRGMHIEALEMYITFGMEDKFSALVLTSLFKMSK 360
 DB 301 NEIAGALKRSQFLVPMVSGIVESSIKRGMHIEALEMYITFGMEDKFSALVLTSLFKMSK 360
 QY 361 ESFERAKGASPLAKFEAAATKQIAVLSSVMQCMETHKLDPAKELPWQIKQIVSLEKD 420
 DB 361 ESFERAKGASPLAKFEAAATKQIAVLSSVMQCMETHKLDPAKELPWQIKQIVSLEKD 420
 QY 421 TLQDKEMEEKARSLMEEAALAKRMYNQIKRPLSPMEMPPTVSSSYSPYRDRSP 480
 DB 421 TLQDKEMEEKARSLMEEAALAKRMYNQIKRPLSPMEMPPTVSSSYSPYRDRSP 480
 QY 481 SQRDDDDQDEISALVSSYLGSTSPHRSRSPSPVWPLPHGGGRSVYAYEHLAPNSYP 540
 DB 481 SQRDDDDQDEISALVSSYLGSTSPHRSRSPSPVWPLPHGGGRSVYAYEHLAPNSYP 540
 QY 541 GGHGHLRHQSPSLVHGQRHPLQYSPPIHGQQQLPYGIQRYRHSPEERYLGLSNQSP 600
 DB 541 GGHGHLRHQSPSLVHGQRHPLQYSPPIHGQQQLPYGIQRYRHSPEERYLGLSNQSP 600
 QY 601 RSNSSLDPK 609
 DB 601 RSNSSLDPK 609

RESULT 2
 ID AAG06057 standard; Protein; 558 AA.
 XX AC
 XX AAG06057;
 XX DT 17-OCT-2000 (first entry)

XX Arabidopsis thaliana protein fragment SEQ ID NO: 2691.
 DE Protein identification; signal transduction pathway; metabolic pathway;
 XX hybridisation assay; genetic mapping; gene expression control; promoter;
 XX termination sequence.
 OS Arabidopsis thaliana.
 PN EP1033405-A2.
 XX 06-SEP-2000.
 XX 25-FEB-2000; 2000EP-0301439.
 XX 25-FEB-1999; 99US-0121825.
 PR 05-MAR-1999; 99US-0123180.
 PR 09-MAR-1999; 99US-0123548.
 PR 23-MAR-1999; 99US-0125788.
 PR 25-MAR-1999; 99US-0128264.
 PR 29-MAR-1999; 99US-0126785.
 PR 01-APR-1999; 99US-0127462.
 PR 06-APR-1999; 99US-0128234.
 PR 08-APR-1999; 99US-0128714.
 PR 16-APR-1999; 99US-0129845.
 PR 19-APR-1999; 99US-0130077.
 PR 21-APR-1999; 99US-0130449.
 PR 23-APR-1999; 99US-0130510.
 PR 23-APR-1999; 99US-0130891.
 PR 28-APR-1999; 99US-0131449.
 PR 30-APR-1999; 99US-0132048.
 PR 30-APR-1999; 99US-0132407.
 PR 04-MAY-1999; 99US-0132484.
 PR 05-MAY-1999; 99US-0132485.
 PR 06-MAY-1999; 99US-0132486.
 PR 07-MAY-1999; 99US-0132487.
 PR 11-MAY-1999; 99US-0132863.
 PR 14-MAY-1999; 99US-0134256.
 PR 14-MAY-1999; 99US-0134218.
 PR 14-MAY-1999; 99US-0134219.
 PR 14-MAY-1999; 99US-0134221.
 PR 14-MAY-1999; 99US-0134370.
 PR 18-MAY-1999; 99US-0134768.
 PR 19-MAY-1999; 99US-0134941.
 PR 20-MAY-1999; 99US-0135124.
 PR 21-MAY-1999; 99US-0135353.
 PR 24-MAY-1999; 99US-0135629.
 PR 25-MAY-1999; 99US-0136021.
 PR 27-MAY-1999; 99US-0136392.
 PR 28-MAY-1999; 99US-0136782.
 PR 01-JUN-1999; 99US-0137222.
 PR 03-JUN-1999; 99US-0137528.
 PR 04-JUN-1999; 99US-0137502.
 PR 07-JUN-1999; 99US-0137724.
 PR 08-JUN-1999; 99US-0138094.
 PR 10-JUN-1999; 99US-0138540.
 PR 10-JUN-1999; 99US-0138847.
 PR 14-JUN-1999; 99US-0139119.
 PR 16-JUN-1999; 99US-0139452.
 PR 17-JUN-1999; 99US-0139453.
 PR 17-JUN-1999; 99US-0139492.
 PR 18-JUN-1999; 99US-0139454.
 PR 18-JUN-1999; 99US-0139455.
 PR 18-JUN-1999; 99US-0139456.
 PR 18-JUN-1999; 99US-0139457.
 PR 18-JUN-1999; 99US-0139458.
 PR 18-JUN-1999; 99US-0139459.
 PR 18-JUN-1999; 99US-0139460.
 PR 18-JUN-1999; 99US-0139461.
 PR 18-JUN-1999; 99US-0139462.
 PR 18-JUN-1999; 99US-0139463.
 PR 18-JUN-1999; 99US-0139750.
 PR 18-JUN-1999; 99US-0139763.

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

3M protein - protein search, using sw model

run on: August 14, 2003, 10:27:44 ; Search time 85 Seconds
(without alignments)
1137.229 Million cell updates/sec

Title: US-09-890-475-1
Perfect score: 3104
Sequence: 1 MSNYPPTVAQAQPTTANPLL.....RYLGLSNQSPRSNSLDPK 609

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_19Jun03.*
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3104	100.0	609	21 AAB08030	Polypeptide which
2	347	11.2	558	21 AAG06057	Arabidopsis thalia
3	346	11.1	548	21 AAG06058	Arabidopsis thalia
4	338	10.9	560	21 AAG39607	Arabidopsis thalia
5	337	10.9	550	21 AAG39608	Arabidopsis thalia
6	336	10.8	454	21 AAG06059	Arabidopsis thalia
7	327	10.5	456	21 AAG39609	Arabidopsis thalia
8	296.5	9.6	532	21 AAG11239	Arabidopsis thalia
9	296.5	9.6	532	21 AAG49607	Arabidopsis thalia

10	292.5	9.4	505	21 AAG11240	Arabidopsis thalia
11	292.5	9.4	505	21 AAG49608	Arabidopsis thalia
12	282	9.1	532	21 AAG19753	Arabidopsis thalia
13	281.5	9.1	483	21 AAG11241	Arabidopsis thalia
14	281.5	9.1	483	21 AAG49609	Arabidopsis thalia
15	277	8.9	505	21 AAG19754	Arabidopsis thalia
16	275	8.9	488	21 AAG19755	Arabidopsis thalia
17	151.5	4.9	2421	22 ABB65643	Drosophila melanog
18	146	4.7	709	23 ABB70293	Human novel polype
19	144	4.6	709	23 ABB70293	Human novel polype
20	144	4.6	725	23 ABB70293	Human hyaluronic a
21	143.5	4.6	1201	20 AAU11436	Human hyaluronic a
22	143	4.6	972	24 ABB19394	Human intracellular
23	142	4.6	725	18 AAU39165	Human RHAMM protei
24	142	4.6	725	23 ABB60842	Human receptor for
25	141	4.5	646	23 ABB70292	Human novel polype
26	140.5	4.5	710	22 ABB78497	Human protein SQ
27	137.5	4.4	2074	21 AAY54319	Amino acid sequenc
28	136.5	4.4	1201	22 ABB58421	Drosophila melanog
29	136	4.4	1143	22 ABB211341	Novel human diagno
30	132.5	4.3	888	23 AAU83013	Human homologue of
31	130.5	4.2	984	22 AAU39331	Human polypeptide
32	130.5	4.2	1084	22 AAU41117	Human polypeptide
33	129	4.2	485	22 AAU39971	Human polypeptide
34	129	4.2	771	22 AAU69462	Human purified sec
35	127.5	4.1	786	20 AAY55937	Human SULU3 protei
36	127.5	4.1	1001	20 AAY55942	Human/Murine SULU3
37	127.5	4.1	1001	23 ABB97326	Novel human protei
38	127.5	4.1	1005	24 ABB47509	Breast cancer asso
39	127.5	4.1	1726	24 AAG16416	Human nucleic acid
40	127.5	4.1	3256	21 AAY50976	Human cell cycle p
41	127.5	4.1	3256	23 ABB77188	Prostate adenocarc
42	127.5	4.1	3256	24 ABB77188	Protein differenti
43	127	4.1	2816	22 AAU07489	Human novel cytoki
44	126	4.1	951	22 AAU68572	Intracellular traf
45	126	4.1	953	22 AAU53070	Intracellular traf

ALIGNMENTS

RESULT 1
AAB08030
ID AAB08030 standard; Protein; 609 AA.
XX
AC AAB08030;
XX
DT 04-DEC-2000 (first entry)
XX
DE Polypeptide which is encoded by the FRI gene of Arabidopsis.
XX
KW H51; one locus-FRIGIDA; FRI gene; flowering time; blotting;
KW flower initiation; stem elongation; flower production.
XX
OS Arabidopsis sp.
XX
PN WO200046358-A2.
XX
PD 10-AUG-2000.
XX
PF 25-JAN-2000; 2000WC-GE00197.
XX
PR 05-FEB-1999; 99GB-0002660.
XX
(PLAN-) PLANT BIOSCIENCE LTD.
XX
Johanson U, West J, Dean C;
XX
WPI; 2000-532899/48.
XX
DR N-PSDB; AAA63669.
XX
PT New nucleic acid derived from the FRI locus of a plant, e.g.
PT Arabidopsis, encoding a polypeptide capable of specifically altering

```

295 AACGATCAAAAGCGCTCTTTGGGAGAGCTTCCAGAAATAATAGAGGAATGTTGGCAAAAC 354
328 GlyMetHisIleGluAlaLeuGluMetValTyrThrPheGlyMetClnuAspLysPheSer 347
355 AAGCAGTAGTAAAGCTGTTTCATATTGTTCTATAGTGTGGAAATGAGGAGATATTAAAC 414
348 AlaAlaLeuValLeuThrSerPheLeuLysMetSerLysGluSerPheGluArgAlaLys 367
415 CCTAGGACACTTTTAAACATTTATAGAGAGAGTGTAAGAAATCAATTCATAACATGAAT 474
368 ArgLysAlaGlnSerProLeuAlaPheLysGluAlaAlaThrLysGlnLeuAlaValLeu 387
475 GGATCAAAAGTGCACATCAAGGAATATTGTAGCAAAAGGGAAGTACTTAGGTGATCTG 534
388 SerSerValMetGlnCysMetGluThrHisLysLeuAspProAlaLysGluLeuProGly 407
535 AGATCGATCATCAATGTTTGAGAAATCATGAGGTGATGATCCATCAAAAGTCTTCCAGGA 594
408 TrpGlnIleLysGluGlnIleValSerLeuGluLysAsp 420
595 TGGGAATAGTAAGAGATAGCGAGCTTGGAGAAAGAA 633

RESULT 15
LOCUS BQ118418 637 bp mRNA linear EST 07-MAR-2003
DEFINITION ESP603994 mixed potato tissues Solanum tuberosum cDNA clone STWEA72
3' end, mRNA sequence.
ACCESSION BQ118418
VERSION BQ118418.1 GI:20170380
KEYWORDS EST.
SOURCE Solanum tuberosum (potato)
ORGANISM Solanum tuberosum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
asterids; lamids; Solanales; Solanaceae; Solanum.
1 (bases 1 to 637)
Buell,C.R., Hart,A., Baker,B., Tanksley,S., Fry,W., Smart,C.,
Reatrepo,S., Griffiths,H., van der Hoeven,R., Teai,J. and
Karamycheva,S.A.
Generation of a set of potato cDNA clones for microarray analyses
Unpublished
Contact: Robin Buell
The Institute for Genomic Research
9712 Medical Center Dr, Rockville, MD 20850, USA
Email: potato-array@tigr.org
This clone can be obtained from the University of Arizona Genomics
Institute. Orders can be made through URL:
http://genome.arizona.edu/orders/
Seq primer: T7.
FEATURES
Location/Qualifiers
1..637
/organism="Solanum tuberosum"
/mol_type="mRNA"
/cultivar="Kennebec or Binjite"
/db_xref="taxon:4113"
/clone="STWEA72"
/tissue_type="mixed tissues"
/lab_host="SOLR"
/clone_lib="mixed potato tissues"
/note="Vector: pBluescript SK(-); Site 1: EcoRI; Site 2:
XhoI; supplier: Combination of untreated and Phytophthora
infestans-treated libraries of stolons, leaves, leaflets,
axillary buds of stem explants, petioles, germinating eyes
, tubers, or roots."
BASE COUNT 180 a 135 c 168 g 154 t
ORIGIN
Alignment Scores:
Pred. No.: 2.85e-24 Length: 637
Score: 371.00 Matches: 82
Percent Similarity: 56.04% Conservative: 34
Best Local Similarity: 39.61% Mismatches: 67
Query Match: 11.95% Indels: 24

```

```

DB: 13 Gaps: 3
US-09-890-475-1 (1-609) x BQ118418 (1-637)
Qy 111 ProMetLeuSerProProArgSerAsnValSerValGluThrThrValThrValSerGln 130
Db 21 CCTATGGAGAGTCTGGCTGGTTCTCCGCTATCCAGACTTCCACCGCCACCGCGCTGCAG 80
Qy 131 Pro---SerGlnGluIleValProGluThrSerAsnLysProGluGlyGlyArg----- 147
Db 81 CCGCAGACGGCTGATACCAACCCCTTCCACCGGCTCACCACCGTCCAGCTCGAATCAAAACC 140
Qy 148 -----MetCysGluLeu 151
Db 141 TCATCCGAAAGAAAAAACAATCTCCGATTCCAGCTCGAAACCCCTTCGAAGACG 200
Qy 152 MetCysSerLysGlyLeuArgLysTyrIleTyrAlaAsnIleSerAspGlnAlaLysLeu 171
Db 201 ATGTGGCAAGGACTTCGAAGTTACATAGTATCGAAACTCGCAGAGAAAAACACACTG 260
Qy 172 MetGluGluIleProSerAlaLeuLysLeuAlaLysGluProAlaLysPheValLeuAsp 191
Db 261 CGCGAAGAACTCACCAAGGCGTTGAAACTTGACCCCAACGACGACAAAGCTTGTACTTAGC 320
Qy 192 CysIleGlyLysPheTyrLeuGlnGlyArgAlaPheThrLysGluSerProMetSer 211
Db 321 TGATGGAGACTTCTTCGAAAGGGCAAGCTTTTGATAGGATGCAGATGATT 380
Qy 212 SerAlaArgGlnValSerLeuLeuIleLeuGluSerPheLeuLeuMetProAspArgGly 231
Db 381 GCCACGAGGAGCATCTGCAITGGTTTGGAGTGTCTTGTGATG----- 428
Qy 232 LysGlyLysValLysIleGluSerTrpIleLysAspGluAlaGluThrAlaAlaValAla 251
Db 429 ---GGTTTCGATGTAATTGATGAAGGGTAAAGGAAGCTGCAAGCAGCAGTGATA 485
Qy 252 TrpArgLysArgLeuMetThrGluGlyGlyLeuAlaAlaAlaGluLysMetAspAlaArg 271
Db 486 TGGAGAAGAGGTTTCTTCATGACGAGGTATAAGAAAGGCTAGTACAAATGGATCGCGG 545
Qy 272 GlyLeuLeuLeuValAlaCysPheGlyValProSerAsnPheArgSerThrAspLeu 291
Db 546 GGATTGCTATTGCTTATTGGGTGTTTGGGATTCACAAATTTTTCATATGAGGACGTT 605
Qy 292 LeuAspLeuIleArgMetSer 298
Db 606 AGGAGCTTGATTCAGTGAGT 626

```

Search completed: August 25, 2003, 13:23:54
Job time : 3528 secs

```

/tissue_type="flower"
/dev_stage="anthesis"
/clone_lib="tomato flower buds, anthesis, Cornell University"
/notes="Vector: pBluescript SK(-); Site 1: EcoRI; Site 2: XhoI; supplier: Tanksley; Flower buds and flowers were taken from greenhouse plants (4-8 wks old, TA496). They were immediately frozen in liquid nitrogen and then size-separated while remaining frozen."
BASE COUNT      192 a  149 c  173 g  171 t
ORIGIN

Alignment Scores:
Pred. No.:      1.58e-25      Length:      685
Score:          385.00      Matches:      94
Percent Similarity: 53.56%      Conservative: 34
Best Local Similarity: 39.33%      Mismatches: 79
Query Match:      12.40%      Indels:      32
DB:              9      Gaps:      6

US-09-890-475-1 (1-609) x AW738032 (1-685)
QY 86 IleGluAsnAlaIleAspSerLysLeuGluSerAsnGlyValValLeuAlaAlaArgAsn 105
DB 35 ATGGAGAACTCGCTGACTCGCGGTATCCAGACTCCAGCGCGCGGG-----85
QY 106 AsnAsnPhelHisGlnProMet-----LeuSerProPro---116
DB 86 -----CAGCAGCGCGCAGCGGTGATACACCCCTTACCGCGGTACACCGGTCA 136
QY 117 ---ArgAsnAsnValSerValGluThrThrValThrValSerGlnProSerGlnGluIle 135
DB 137 GCCTCGAATCAAACTCATCCGAGAAAAGACTCAAAATCTCCGATTCAAG-----190
QY 136 ValProGluThrSerAsnLysProGluGlyGlyArgMetCysGluLeuMetCysSerLys 155
DB 191 ---CTCGAAACC-----CTCTGCAAGCAGCATGTGCGGCAAG 223
QY 156 GlyLeuArgLysTyrlleTyAlaAsnIleSerAspGlnAlaLysLeuMetGluGluIle 175
DB 224 GGACTTCGAAGTACATATTATCGAACTCCAGAGAAAACACACTCGCGCAAGAACTC 283
QY 176 ProSerAlaLeuLysLeuAlaLysGluProAlaLysPheValLeuAspCysIleGlyLys 195
DB 284 ACCAAGGCATTGAACCTTGACCCCAATGAGCAAGCAAGCTTGTTAGCTGTGTGGAGAC 343
QY 196 PheTyrlleGlnGlyArgAlaPheThrLysGluSerProMetSerSerAlaArgGln 215
DB 344 TTTCTTCTAAAGCGCAAGCTTTTGATAGGATGCTCGATGATGCCACAGGGA 403
QY 216 ValSerLeuLeuIleLeuGluSerPheLeuLeuMetProAspArgGlyLysVal 235
DB 404 GCATCTGCATTGTTTGGAGTCTTTCTTGTGATA-----GGTTTCGAT 448
QY 236 LysIleLeuSerTrpIleLysAspGluAlaGluThrAlaAlaValAlaTrpArgLysArg 255
DB 449 GTAATTGTAAGCGGTAAAGGAAGAGCTGCACAGCAGCAGTATATGGAGAAGAG 508
QY 256 LeuMetThrGluGlyLysLeuAlaAlaGluLysMetAspAlaArgGlyLeuLeu 275
DB 509 TTTGTTGATCAACAGAGTATACAAAGGCTAGTACAAATCGATGCTCGGGATTGCTATTG 568
QY 276 LeuValAlaCysPheGlyValProSerAsnPheArgSerThrAspLeuAspLeuIle 295
DB 569 CTTATTGGGTGTTTGGGATTCCTCCCAATTTTACATGAGACATTAGGACTTGATT 628
QY 296 ArgMetSerGlySerAsnGluIleAlaGlyAlaLeuLysArgSerGlnPheLeuVal 314
DB 629 CGACTGAGTATATCAGGAGAGATATCTACTGCCTCAGGAGATCAATGTGCTTATG 685

```

```

RESULT 14
AW738612      651 bp      mRNA      linear      EST 18-MAY-2001
LOCUS

```

```

DEFINITION EST340039 tomato flower buds, anthesis, Cornell University
ACCESSION AW738612
VERSION AW738612.1 GI:7647557
KEYWORDS EST.
SOURCE Lycopersicon esculentum (tomato)
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Eudicotyledons; core eudicots;
asterids; lamids; Solanales; Solanaceae; Solanum; Lycopersicon.
REFERENCE 1 (bases 1 to 651)
AUTHORS van der Hoeven,R.S., Bezzerides,J.L., Matern,A.L., Holt,I.E., Liang
,F., Hansen,T.S., Craven,M.B., Bowman,C.L., Renning,C.M., Nierman
,W., Fraser,C.M., Martin,G.B., Giovannoni,J.J. and Tanksley,S.D.
TITLE Generation of ESTs from tomato flower tissue, anthesis
JOURNAL Unpublished
COMMENT Contact: CUGI
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: http://www.genome.clemson.edu/orders/index.html
5 prime sequence.
FEATURES
source
1..651
/organism="Lycopersicon esculentum"
/mol_type="mRNA"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="CTOD7N12"
/tissue_type="flower"
/dev_stage="anthesis"
/clone_lib="tomato flower buds, anthesis, Cornell University"
/notes="Vector: pBluescript SK(-); Site 1: EcoRI; Site 2:
XhoI; supplier: Tanksley; Flower buds and flowers were
taken from greenhouse plants (4-8 wks old, TA496). They
were immediately frozen in liquid nitrogen and then
size-separated while remaining frozen."
BASE COUNT      214 a   89 c   170 g   178 t
ORIGIN

```

```

Alignment Scores:
Pred. No.:      1.92e-24      Length:      651
Score:          373.00      Matches:      74
Percent Similarity: 61.03%      Conservative: 56
Best Local Similarity: 34.74%      Mismatches: 81
Query Match:      12.02%      Indels:      2
DB:              9      Gaps:      1

```

```

US-09-890-475-1 (1-609) x AW738612 (1-651)
QY 208 SerProMetSerSerAlaArgGlnValSerLeuLeuIleLeuGluSerPheLeuLeuMet 227
DB 1 GCACCTTTGGGTCGGGGAAGGAGGCTAAAGTATTGATTTGGAGTCTGTTGTTAATG 60
QY 228 ProAspArgGlyLysGlyLysValLysIleGluSerTrpIleLysAspGluAlaGluThr 247
DB 61 ATTGGAACTACGATAAGGAAATCGAGCTTGAGAAATGGGTGAAGGAAGCGCGGAGCAT 120
QY 248 AlaAlaValAlaTrpArgLysArgLeuMetThrGluGlyGlyLeuAlaAlaGluLys 267
DB 121 ACAGCTTTAGCATGTTGGAGAGGATGAATCTCAAGGAGGA-----AGCCTCAAGAA 174
QY 268 MetAspAlaArgGlyLeuLeuLeuValAlaCysPheGlyValProSerAsnPheArg 287
DB 175 ATCGATGCCCGAGGTTTCTCTTTATTGGGGTTTGGGATTCAGATCAATTTACA 234
QY 288 SerThrAspLeuAspLeuIleArgMetSerGlySerAsnGluIleAlaGlyValLeu 307
DB 235 AATGCGAATATCAGAAATTTGTTTCAGGTAAGTAACCTTAAGTCGAGTGTGATGCACTC 294
QY 308 LysArgSerGlnPheLeuValProMetValSerGlyIleValGluSerSerIleLysArg 327
DB 308 LysArgSerGlnPheLeuValProMetValSerGlyIleValGluSerSerIleLysArg 327

```

```

2Y 159 LysTyrIleTyrAlaAenIleSerAspGlnAlaLysLeuMetGluGluIleProSerAla 178
Db 521 AGATACATGACTCGATATATCTGTGACCGAGCTTAAGCTGATTGAAGAGCTTCTGCAGCT 580

2Y 179 LeuLysLeuAlaLys-GluProAlaLys-PheValLeuAspCys-IleGlyLysPheTyr 197
Db 581 CTGAAGTACCAAGGAGGAGCAGTAAGTNCGTGTGGAAATGCANTGCGAAGTTTTC 640

Y 198 LeuGlnGlyArgAlaPheThrLysGluSerProMetSerSerAla 213
Db 641 TTACAGGGGGCAAAGCTTATGGAGTGAATCCCATATGATCCCTCGC 688

RESULT 12
LOCUS BG890481 671 bp mRNA linear EST 07-MAR-2003
DEFINITION EST16332 cSTD Solanum tuberosum cDNA clone cSTD1821 5' sequence,
mRNA sequence.
ACCESSION BG890481
VERSION BG890481.1 GI:14267593
KEYWORDS EST.
SOURCE Solanum tuberosum (potato)
ORGANISM Solanum tuberosum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
asterids; lamiales; Solanales; Solanaceae; Solanum.
REFERENCE 1 (bases 1 to 671)
AUTHORS van der Hoeven,R., Bezzerides,J., Ewing,E., Cho,J., Chieming,A.,
Bougri,O., Buell,C.R., Renning,C., Tanksley,S. and Baker,B.
Generation of ESTs from dormant potato tubers
TITLE Unpublished
JOURNAL
COMMENT Contact: Robin Buell
The Institute for Genomic Research
9712 Medical Center Dr, Rockville, MD 20850, USA
Email: potato-array@tigr.org
This clone can be obtained from the University of Arizona Genomics
Institute. Orders can be made through URL:
http://genome.arizona.edu/orders/
Seq primer: M13F-R.
FEATURES
source
location/Qualifiers
1..671
/organism="Solanum tuberosum"
/mol_type="mRNA"
/cultivar="Kennebec"
/db_xref="taxon:4113"
/clone="cSTD1821"
/tissue_type="dormant tuber"
/dev_stage="one month post-harvest"
/lab_host="SOLR"
/clone_lib="cSTD"
/notes="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI; This library targets genes expressed in dormant
tubers. This library was made from sections of dormant
tubers, avoiding the buds and epidermis. Tubers were stored
for one month post-harvest at 40C. The tuber was peeled,
well away from the surface. Then it was chopped into 1-2
mm cubes and immediately frozen in liquid nitrogen. This
library is noted as P4 in Tanksley lab notebooks."
BASE COUNT 193 a 110 c 178 g 190 t
ORIGIN
Alignment Scores:
Pred. No.: 4.05e-29 Length: 671
Score: 423.50 Matches: 90
Percent Similarity: 65.40% Conservative: 48
Best Local Similarity: 42.65% Mismatches: 70
Query Match: 13.64% Indels: 3
Gaps: 12
JB: 12

JS-09-890-475-1 (1-609) x BG890481 (1-671)

2Y 156 GlyLeuArgLysTyrIleTyrAlaAsnIleSerAspGlnAlaLysLeuMetGluGluIle 175
Db 3 GGTCTGAAGAAGTACATGGTAAAGCATCTTCGGATATAAAAAAAGACTGCTGCGAAGATGC 62

```

```

176 ProSerAlaLeuLysLeuAlaLysGluProAlaLysPheValLeuAspCysIleGlyLys 195
Db 63 CCCAAGGCATTTGAAATCTCGGAAATCCAGCAATGCTCGTGTGTGAAATGTCTGCACAAAG 122

196 PheTyrLeuGlnGlyArgAlaPheThrLysGluSerProMetSerSerAlaArgGln 215
Db 123 TTTTATCTTCAGGGAGCCGTCCTACGTTCAAGGGTTCACATATGATCCTTCAAAGGAA 182

216 ValSerLeuLeuLeuGluSerPheLeuLeuMetProAspArgGlyLysGlyLysVal 235
Db 183 GCTTCTGTATTTGGCTTTGGAGTGCTTATTTGTCAGCATGGTAGGGGAGAGAGTAGTT 242

236 LysIleLeuSerTrpIleLysAspGluAlaGluThrAlaAlaValAlaIleTrpArgLysArg 255
Db 243 GAGATTCGGAGGGGTGAAGAAAGATGCCAGCAGGAGCTCTAGCATGGAGAGAGG 302

256 LeuMetThrGluGlyLeuAlaAlaGluLysMetAspAlaArgGlyLeuLeuVal 275
Db 303 TTGAATTCGGAGGAGGGGTAAAGAAAGCTTATGACATGGATGCCGAGGTTTGTCTATTG 362

276 LeuValAlaCysPheGlyValProSer---AsnPheArgSerThrAspLeuLeuAspLeu 294
Db 363 CTCTCTGGATGTTTGGTATTCATCATATGATGATTTTACATTCGGATATCATGTTATTG 422

295 IleArgMetSerGlySerAsnGluIleAlaGlyAlaLeuLysArgSerGlnPheLeuVal 314
Db 423 CTTCAGTTG--GGCGTCAAGTTAAATTTCTGTGCTCTCAGGAGATCAAGTATTTTTCATG 479

315 ProMetValSerGlyIleValGluSerSerIleLysArgGlyMetHisIleGluAlaLeu 334
Db 480 GAATATATCCCAAACTAATACAGTGGGGGGTGAAG--GGCAATGTAATTAACCAATT 536

335 GluMetValTyrThrPheGlyMetGluAspLysPheSerAlaAlaLeuValLeuThrSer 354
Db 537 GACATTCCTATATCTTTGTTTGGTAAAGAACTTGACCTCTCGGACTCTTAACATCA 596

355 PheLeuLysMetSerLysGluSerPheGluArg 365
Db 597 TTTTAAAGAGTCCGAGGAGACATTTGTCGAAA 629

RESULT 13
LOCUS AW738032 685 bp mRNA linear EST 18-MAY-2001
DEFINITION EST39459 tomato flower buds, anthesis, Cornell University
Lycopersicon esculentum cDNA clone cTOD5C11.5', mRNA sequence.
ACCESSION AW738032
VERSION AW738032.1 GI:7646977
KEYWORDS EST.
SOURCE Lycopersicon esculentum (tomato)
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
asterids; lamiales; Solanales; Solanaceae; Solanum; Lycopersicon.
REFERENCE 1 (bases 1 to 685)
AUTHORS van der Hoeven,R.S., Bezzerides,J.L., Matern,A.L., Holt,I.E., Liang
,F., Hansen,T.S., Craven,M.B., Bowman,C.L., Renning,C.M., Nierman
,W., Fraser,C.M., Martin,G.B., Giovannoni,J.J. and Tanksley,S.D.
Generation of ESTs from tomato flower tissue, anthesis
TITLE Unpublished
JOURNAL
COMMENT Contact: CJGJ
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: http://www.genome.clemson.edu/orders/index.html
5 prime sequence.
location/Qualifiers
1..685
/organism="Lycopersicon esculentum"
/mol_type="mRNA"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="cTOD5C11"

```

```

Pred. No.: 1.45e-30 Length: 771
Score: 440.00 Matches: 99
Percent Similarity: 56.35% Conservative: 43
Best Local Similarity: 39.29% Mismatches: 87
Query Match: 14.18% Indels: 24
DB: 12 Gaps: 3

US-09-890-475-1 (1-609) x B1176101 (1-771)
Qy 111 ProMetLeuSerProProArgAsnAsnValSerValGluThrThrValThrValSerGln 130
Db 36 CCCAGACTCCACCGCCACCGCCGCGCAG-ACGGCTGATACCACC----- 79
Qy 131 ProSerGlnGluLeuValProGluThrSerAsnLysProGluGlyGlyArg----- 147
Db 80 CTTTACCGCGCTCACCACTGAGCTGATCAACCTCATCCGAGAAAGAACTCA 139
Qy 148 -----MetCysGluLeuMetCysSerLysGlyLeuArg 158
Db 140 AAATCTCCGATTACAGCTCGAAACCTCTGCAAGACGATGCGGCAAGGACTTCCA 199
Qy 159 LysTyrIleThrAlaSerAspGlnAlaLysLeuMetGluGluIleProSerAla 178
Db 200 AGTTACATAGTATCGAATCCAGAGAAACACACTGCGGAGAACTCACCAAGCA 259
Qy 179 LeuLysLeuAlaLysGluProAlaLysPheValLeuAspCysIleGlyLysPheTyrLeu 198
Db 260 TTGGAACCTTCAACCAATGAGCAAGCTGTACTTAGCTGTATGGAGACTTCTTTCG 319
Qy 199 GlnGlyArgAlaPheThrLysGluSerProMetSerSerAlaArgGlnValSerLeu 218
Db 320 AAAGGGCAAGACTTTGATAGATGACAGATGATTCACCAAGGAGCACTCTCCA 379
Qy 219 LeuIleLeuGluSerPheLeuLeuMetProAspArgGlyLysGlyLysValLysIleGlu 238
Db 380 TTGTTTGGAGTGTCTTCTTGTGATG-----GGTTTCATGATGATTCAT 424
Qy 239 SerTrpIleLysAspGluAlaGluThrAlaAlaValAlaIleTrpArgLysArgLeuMetThr 258
Db 425 GAAGGGGTAAGGAAGAGCTGCACAGCAGCAGTAAATATGAGAGAAGGTTTGTGAT 484
Qy 259 GluGlyLysLeuAlaAlaGluLysMetAspAlaArgGlyLeuLeuLeuValAla 278
Db 485 GAACGAGTATAGAAGGCTGTACATGGATGCGCGGGATTCCTATTGCTATTGGG 544
Qy 279 CysPheGlyValProSerAsnPheArgSerThrAspLeuLeuAspIleArgMetSer 298
Db 545 TGTTTGGGATTCACAAATTTTTCATAGAGACGTTTAGGACCTTGTATCGAGTGAGT 604
Qy 299 GlySerAsnGluIleAlaGlyAlaLeuLysArgSerGlnPheLeuValProMetValSer 318
Db 605 AATATCAGGAGATTTCTACTGCACTCAGGAGATCAATGTGCTATGGAAGATTCCA 664
Qy 319 GlyIleValGluSerSerIleLysArgGlyMetHis-IleGluAlaLeuGluMetValTy 338
Db 665 GAAATAATAGAGGCAATGTTGTAAGCATAAGATGGAAGTTGATGCTGTAATATGTTGCCTA 724
Qy 338 rThrPheGlyMetGluAspLysPheSerAlaAla 349
Db 725 TACTTTTGGAGTTAGGAAAAATGATAGCCCTTGC 758

RESULT 11
BH984132
LOCUS cd42a06.b1 B.oleracea002 Brassica oleracea genomic, genomic survey
DEFINITION
ACCESSION BH984132
VERSION BH984132.1 GI:23516303
KEYWORDS
SOURCE Brassica oleracea
ORGANISM Brassica oleracea
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids

```

```

; eurosids II; Brassicales; Brassicaceae; Brassica.
1 (bases 1 to 688)
Delehaunty,K., Fewell,G., Fulton,L., McCombie,W.R., Miner,T., Nash
W., Rabinowicz,P.D. and Wilson,R.K.
Whole genome shotgun reads from Brassica oleracea
Unpublished
Contact: Richard K. Wilson
Genome Sequencing Center
Washington University School of Medicine
Email: submissions@watson.wustl.edu
Plate: odf42 row: a column: 06
Seq primer: -21UPPOT forward
Class: shotgun
High quality sequence start: 103
High quality sequence stop: 551.
Location/Qualifiers
1. .688
/organism="Brassica oleracea"
/mol_type="genomic DNA"
/db_xref="taxon:3712"
/clone_lib="B.oleracea002"
/note="Vector: pOTw13; Whole genome shotgun library from
flowering buds. DNA was purified from a crude nuclear
prep using Brassica oleracea TOL000DH3 buds provided by
Thomas Osborn at the University of Wisconsin. Genomic
DNA was provided by Pablo Rabinowicz (CSHL) and the
shotgun library prepared at Washington University Genome
Sequencing Center."
BASE COUNT 179 a 192 c 155 g 160 t 2 others
ORIGIN
Alignment Scores: 3.22e-30 Length: 688
Pred. No.: 435.50 Matches: 102
Score: 63.30% Conservative: 36
Percent Similarity: 46.73% Mismatches: 46
Best Local Similarity: 14.03% Indels: 34
Query Match: 28 Gaps: 5
DB:
US-09-890-475-1 (1-609) x BH984132 (1-688)
Qy 9 AlaAlaGlnProThrThrThrAlaAsn-----ProLeuLeuGlnArgHisGln 24
Db 104 GCCCGTGTATCCATCCACGAGGAGAAATAAACCATCATCGCGGACCATACAACGG----- 157
Qy 25 SerGluGlnArgArgArgGluLeuProLysIleValGluThrGluSerThrSerMetAsp 44
Db 158 -----GGAACCGTGCCTACAAACACCGAA 181
Qy 45 IleThrIleGlyGlnSerLysGlnProGlnPheLeuLysSerIleAspGluLeuAlaAla 64
Db 182 ATCAGATCGAACAATCTAACCATCGCAATTTTGAATCGATCGACGATTTAACTGCG 241
Qy 65 PheSerValAlaValGluThrPheLysArgGlnPheAspAspLeuGlnLysHisIleGlu 84
Db 242 TTTGCAGCTCGCATGCGCCTTCAACGCCCACTACGACGACTTGCAAAACCACTGGAT 301
Qy 85 SerIleGluAsnAlaIleAspSerLysLeuGluSerAsnGlyValValLeuAlaAlaArg 104
Db 302 TACATCAAGAAGCCCATTTGGCTCCAGTCTCAAAATTCAAAGGCATCATCGCGAGTCTCC 361
Qy 105 AsnAsnAsnPheHisGlnProMetLeuSerProProArgAsnAsnValSerValGluThr 124
Db 362 TCCTCCCGATCGCAGTCT-----CCACGGAACGATGCTTCGCGAGAACA 406
Qy 125 ThrValThrValSer-----GlnProSerGlnGluIleValProGlu 138
Db 407 GCCACCGCGGTTGCCGCCACACATCGCGGCCCAAGGAGACTTCTGACAGACTACCGGAG 466
Qy 139 ThrSerAsnLysProGluGlyArgMetCysGluLeuMetCysSerLysGlyLeuArg 158
Db 467 ATTTGGATTAAGGTGGAG-----CGATTGCGAGTTGATGTGCGAGCAAGGCGCTGCGT 520

```


Db	183	GAGGA-----GTTGAGTGAAGAAGAGTGAAGAAGAGGCGGAGAGGCGAGCT	233
QY	250	ValAlaTTPArgLysArgLeuMetThrGlucGlyLeuAlaLaalGluLysMetAsp	269
Db	234	TTAGCATGCAGAGAGAGGTTGATGCTGTAAGGAGGTTTACGAAAGGTGATGATATGAT	293
QY	270	AlaArgGlyLeuLeuValAlaCysPheGlyValProSerAsnPheArgSerThr	289
Db	294	GCCCGGGTTGTCATGCTTATTGGGTGTTTCGGATTCCAGAGGAGGATTCAGTAAATGAG	353
QY	290	AspLeuLeuAspLeuLeuArgMetSerGly---SerAsnGluLeuAlaGlyAlaLeuLys	308
Db	354	GATATCAGGAGTATTATTTCTGCGAAGTCCGTTCAAGAAGATAATTTCTTGTCCTCACA	413
QY	309	ArgSerGlnPheLeuValProMetValSerGlyIleValGluSerSerIleLysArgGly	328
Db	414	AGATCAAAATGCTTCATGCGCAAGATTATACAGAAATATAGAGGGGATGGTGACCAAG	473
QY	329	MethIleGluAlaLeuGluMetValThrPheGlyMetGluAspLysPheSerAla	346
Db	474	ATGGAAATGGAGGAGTGTATCTTCCTATATCTTTGGAATGGAGGACAGATTAACTCT	533
QY	349	AlaLeuValLeuThrSerPheLeu	356
Db	534	CAGAAACTGTGACACATATTTA	557
RESULT 10			
LOCUS	BI176101	771 bp mRNA linear EST 07-MAR-2003	
DEFINITION	EST517184 cSTS Solanum tuberosum cDNA clone cSTS24N12 5' sequence, mRNA sequence.		
ACCESSION	BI176101		
VERSION	BI176101.1	GI:14641912	
KEYWORDS	EST		
SOURCE	Solanum tuberosum (potato)		
ORGANISM	Solanum tuberosum		
REFERENCE	1 (bases 1 to 771)		
AUTHORS	van der Hoeven, R., Bezzerides, J., Sun, H., Cho, J., Chiemiango, A., Bougri, O., Buell, C.R., Ranning, C., Tanksley, S. and Baker, B.		
TITLE	Generations of ESTs from sprouting potato eyes		
JOURNAL	Unpublished		
COMMENT	Contact: Robin Buell The Institute for Genomic Research 9712 Medical Center Dr, Rockville, MD 20850, USA Email: potato-array@tigr.org This clone can be obtained from the University of Arizona Genomics Institute. Orders can be made through URL: http://genome.arizona.edu/orders/ Seq primer: M13P-R.		
FEATURES	Location/Qualifiers		
source	1..771		
	/organism="Solanum tuberosum"		
	/mol_type="mRNA"		
	/cultivar="Kennebec"		
	/db_xref="taxon:4113"		
	/clone="cSTS24N12"		
	/tissue_type="sprouting eyes from tubers"		
	/dev_stage="12-14 weeks post harvest"		
	/lab_host="SOLR"		
	/clone_lib="cSTS"		
	/note="Vector: pBluescript SK(-); Site 1: EcoRI; Site 2: XhoI; Various sizes of sprouting eyes (2mm to 15mm) were taken from tubers. The tubers were incubated at 26C in the dark for 2-3 weeks prior to sprouting. The eyes were frozen in liquid nitrogen immediately upon removal from tubers."		
BASE COUNT	229 a	154 c	200 g
ORIGIN			188 t
Alignment Scores:			

Pred. No.: 3,52e-35 Length: 520
 Score: 487.00 Matches: 101
 Percent Similarity: 76.87% Conservative: 12
 Best Local Similarity: 68.71% Mismatches: 28
 Query Match: 15.69% Indels: 6
 DB: 29 Gaps: 3

US-09-890-475-1 (1-609) x BZ512352 (1-520)

Qy 377 LysGluAlaAlaThrLysGlnLeuAlaValLeuSerSerValMetGlnCysMetGluThr 396
 Db 91 AAGAGGCCAACCAAAAGTTTGTAGCTGCTTGTATCATGATGAAGTGTGGAGGCT 150
 Qy 397 HisLysLeuAspProAlaLysGluLeuProGlyTrpGlnIleLysGluGlnIleValSer 416
 Db 151 CACAACCTTAGACCAGAGAAGATGTAACAAGATGTCAGATCAAGCAAAATGATTAG 210
 Qy 417 LeuGluLysAspThrLeuGlnLeuAspLysGluMetGluLysAlaArgSerLeuSer 436
 Db 211 TTGGAGAAGACATCTTCACTGACAAACAGATGCAAGGGGAAGACATCCATCAGT 270
 Qy 437 LeuMetGluGluAlaAlaLeuAlaLysArgMetTyAsnGlnIleLysArgProArg 456
 Db 271 TTAATGGAGGAACAGCATTTGACAGAGAGATTGTATAACCAACAGATGAACGTCCAAG 330
 Qy 457 LeuSerProMetGluMetProProValThrSerSerSerTySerProIleTyArgAsp 476
 Db 331 TTGTCAGACATGGAAATGCCACAGCAGCTTCTCATCTTATCTTCTACCTACCTGCC 390
 Qy 477 ArgSerPheProSerGlnArgAspAspGluAspGluLysSerAlaLeuValSerSer 496
 Db 391 CGAAGCTTCCCTAGTCACAGAC-----GATGAATATCATGCTTCTGCTAGTAGT 441
 Qy 497 TyrLeuGlyProSerThrSerPheProHisArgSer-----ArgArgSerProGluTyr 514
 Db 442 TACTCTGAGCCTTCCACAGGTTTCTCATCGTCAAGTCTCAGGAGTTCCTCTGATAT 501
 Qy 515 MetValProLeuProHisGly 521
 Db 502 TTAGCTCCA---CCTAGTGGT 519

RESULT 8
 BQ115269
 LOCUS ES7600845 mixed potato tissues Solanum tuberosum cDNA clone STMZ77
 DEFINITION 5', end, mRNA sequence.
 ACCESSION BQ115269
 VERSION BQ115269.2 GI:21916872
 KEYWORDS EST
 SOURCE Solanum tuberosum (potato)
 ORGANISM Solanum tuberosum

REFERENCE 1 (bases 1 to 785)
 AUTHORS Buell C.R., Hart A.A., Baker B., Tankalev S., Fry W., Smart C.,
 Restrepo S., Griffiths H., van der Hoeven R., Tsai J. and
 Karamycheva S.A.

TITLE Generation of a set of potato cDNA clones for microarray analyses
 JOURNAL Unpublished
 COMMENT On Apr 17, 2002 this sequence version replaced gi:20167231.

Contact: Robin Buell
 The Institute for Genomic Research
 9712 Medical Center Dr, Rockville, MD 20850, USA
 Email: potato-array@tigr.org
 This clone can be obtained from the University of Arizona Genomics
 Institute. Orders can be made through URL:
 http://genome.arizona.edu/orders/
 Seq primer: T3.

FEATURES
 source
 1..785 Location/Qualifiers
 /organism="Solanum tuberosum"
 /mol_type="mRNA"

/cultivar="Kennebec or Binjete"
 /db_xref="taxon:4113"
 /clone="STMZ77"
 /issue_type="mixed tissues"
 /lab_host="SOLR"
 /clone_lib="mixed potato tissues"
 /note="Vector: pBluescript SK(-); Site 1: EcoRI; Site 2:
 XhoI; supplier: Combination of untreated and Phytophthora
 infestans-treated libraries of stolons, leaves, leaflets,
 axillary buds of stem explants, petioles, germinating eyes,
 tubers, or roots."
 BASE COUNT 224 a 134 c 209 t
 ORIGIN

Alignment Scores:
 Pred. No.: 7,08e-35 Length: 785
 Score: 486.50 Matches: 108
 Percent Similarity: 62.90% Conservative: 48
 Best Local Similarity: 43.55% Mismatches: 74
 Query Match: 15.67% Indels: 18
 DB: 13 Gaps: 5

US-09-890-475-1 (1-609) x BQ115269 (1-785)

Qy 104 ArgAsnAsnAspPheHisGlnProMetLeuSerProProArgAsnVal--SerValG 123
 Db 69 AAGAACAGAACCATCAT-----CGGAATTCATCCCTCTGAAG 107
 Qy 123 LuThrThrValThrValSerGlnProSerGlnGluIleValProGluThrSerAsnLysP 143
 Db 108 AAGAAGAGTGGAGTGAATCCCTTGTGTAGACTA-----AAATCTACTCGTT 158
 Qy 143 roGluGlyGlyArgMetCysGluLeuMetCysSerLysGlyLeuArgLysTyrIleTyrA 163
 Db 159 CAGAGCTAGAAAGCATCTCTGAAAGGACGACGCGTCTGGGAGTACATGATAA 218
 Qy 163 laAsnIleSerAspGlnAlaLysLeuMetGluGluIleProSerAlaLeuLysLeuAlaL 183
 Db 219 CGCATCTCTGATATAAATGACTGCTTGAGGAAGTCCCTAAGGCATTGAACACTCTGC 278
 Qy 183 ysGluProAlaLysPheValLeuAspCysIleGlyLysPheTyrLeuGlnGlyArgArgA 203
 Db 279 GCAATCGGCAAGCTTGATTGGATTGTGGAGAGTTTATTGCAAGGAGGACGG 338
 Qy 203 laPheThrLysGluSerProMetSerSerAlaArgGlnValSerLeuLeuLeuGluS 223
 Db 339 CATATGTTAAGGTTTCACTCCGCTCAATGGAAGGAGCTTCTATATGTTGGTTTGGATT 398
 Qy 223 erPheLeuLeuMetPro--AspArgGlyLysGlyLysValLysIleGluSerTrpIleL 242
 Db 399 GCTTCTTGTTGATGGAAATCGACAGGGA-----GTTGAGTTTGAGAGAGGTGA 449
 Qy 242 ysAspGluAlaGluThrAlaAlaValAlaTyrArgLysArgLeuMetThrGluGlyGlyL 262
 Db 450 AAGAAGAGTCCGGAAGGAGGAGCTTTAGCATGGAAGAGAGGTGATTGCTGAAGGAGTT 509
 Qy 262 euAlaAlaAlaGluLysMetAspAlaArgGlyLeuLeuLeuValAlaCysPheGlyV 282
 Db 510 TTGCAAGGCGTATGATATGATATGCCGGGGTTGCTATTGTTAATTTGGTGTTTCGGA 569
 Qy 282 alProSerAsnPheArgSerThrAspLeuLeuAspLeuIleArgMetSerGly---SerA 301
 Db 570 TTCCAGGAGAAATTTAGAAATAGGATATCAGGAGTTTGTCTCTGCCAGTCCGTTCAAGA 629
 Qy 301 snGluIleAlaGlyAlaLeuLysArgSerGlnPheLeuValProMetValSerGlyIleV 321
 Db 630 AGAATATTTCTGTTCCCTCACAGATCAAAATGCTTTCATGGGCAAGATTACAGAAATAA 689
 Qy 321 alGluSerSerIleLysArgGlyMetHisIleGluAlaLeuGluMetValTyrThrPhe- 340
 Db 690 TAGAGGAGATGGTGAATCAGAAGATGGGAATGAGCAGTTCATATGCTATATCTTGTG 749
 Qy 341 GlyMetGluAspLysPhe 346

```

Db      428 TTTCTGATCTTACCATGTTTGGCTTTTGTCTTCTTACAGAAAGAGGCGCCACCAAG 369
Qy      383 GlnLeuAlaValLeuSerSerValMetGlnCysMetGluThrHisLysLeuAspProAla 402
Db      368 TTTTATAGTGGCTTGTATCAGTCATGAGTGTGTGGAGGCTCACAACTTAGACCCAGAG 309
Qy      403 LysGluLeuProGlyTrpGlnIleLysGluGlnIleValSerLeuGluLysAspThrLeu 422
Db      308 AAAGAAGTACAAGATGGCAGATCAAGACCAATGATTAAAGTTGGAGAAGACATCTT 249
Qy      423 GlnLeuAspLysGluMetGluGluLysAlaArgSerLeuSerLeuMetGluLysAlaLala 442
Db      248 CAACCTCGAACAACAGATCAAGGGAAGCAAGATCCATCAGTTTAAATGGAGGAACAGCA 189
Qy      443 LeuAlaLysArgMetTyrAsnGlnGlnIleLysArgProArgLeuSerProMetGluMet 462
Db      188 TTGACGAAGAGATTGTATPAACACACATGATGAAGCTCCAAAGTTGTGACATGGAATG 129
Qy      463 ProProValThrSerSerSerTyrSerProIleTyrArgAspArgSerPheProSerGln 482
Db      128 CCACGACGAGCTTCTCATCTTATCTTACCTACCTGCGGAGCTTCCCTAGTCAC 69
Qy      483 ArgAspAspGlnAspGlnIleSerAlaLeuValSerSerTyrLeuGlyProSerThr 502
Db      68 AGAGAC-----GATGAATATCAGCTCTTGTCAAGTAGTTACCTCGAGGCTTCACCA 18
Qy      503 SerPheProHisArg 507
Db      17 GGTTCCTCATCGG 3

```

```

RESULT 6
B77833/c
LOCUS   B77833 635 bp DNA linear GSS 16-JAN-1998
DEFINITION T29B23TR TAMU Arabidopsis thaliana genomic clone T29B23, genomic survey sequence.
ACCESSION B77833.1 GI:2774472
VERSION B77833.1
KEYWORDS Arabidopsis thaliana (thale cress)
SOURCE Arabidopsis thaliana
ORGANISM Arabidopsis thaliana
REFERENCE 1 (bases 1 to 635)
AUTHORS Rounsley,S.D., Field,C.B., Bass,S., Linher,K., Linher,K., Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Adams,M.D. and Venter,J.C.
TITLE A BAC End Sequence Database for Identifying Minimal Overlaps in Arabidopsis Genomic Sequencing. Update 3
JOURNAL Unpublished
COMMENT Other_GSSs: T29B23TF
Contact: Steve Rounsley
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: rounsley@tigr.org
Seq primer: M13 Reverse
Class: BAC ends
High quality sequence stop: 635.
Location/Qualifiers
1. .635
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="T29B23"
/sex="hermaphrodite"
/clone_lib="TAMU"
/note="Vector: BclOBACII; Site_1: HindIII; Site_2: HindIII
; Produced by Rod Wing"
209 a 110 c 143 g 173 t

```

FEATURES

```

source
1. .635
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="T29B23"
/sex="hermaphrodite"
/clone_lib="TAMU"
/note="Vector: BclOBACII; Site_1: HindIII; Site_2: HindIII
; Produced by Rod Wing"
209 a 110 c 143 g 173 t

```

BASE COUNT

ORIGIN

```

Alignment Scores: 2.93e-40 Length: 635
Pred. No.: 543.00 Matches: 98
Score: 100.00% Conservative: 0
Percent Similarity: 100.00% Mismatches: 0
Best Local Similarity: 17.49% Indels: 0
Query Match: 28 Gaps: 0
DB:

```

US-09-890-475-1 (1-609) x B77833 (1-635)

```

Qy      512 ProGluTyrMetValProLeuProHisGlyGlyLeuGlyArgSerValTyrAlaTyrGlu 531
Db      629 CCGGATATATGTTCCATTCACATGTCGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 570
Qy      532 HisLeuAlaProAsnSerTyrSerProGlyHisGlyHisArgLeuHisArgGlnTyrSer 551
Db      569 CATCTGCCCCCAATTCATATTTCCAGGTCCAGGTCCAGGTCCAGGTCCAGGTCCAGGTCT 510
Qy      552 ProSerLeuValHisGlyGlnArgHisProLeuGlnTyrSerProIleHisGlyGln 571
Db      509 CCGTCTTTTGGTTCAGGACAGACATCCACTACAGTACTCTCTCTCAATTTCATGGACAA 450
Qy      572 GlnGlnLeuProTyrGlyIleGlnArgValTyrArgHisSerProSerGluGluArgTyr 591
Db      449 CAACAGTTACCATATGTTATACAAAGGTTTACAGATTCACCATCTCGAAGAAAGATAT 390
Qy      592 LeuGlyLeuSerAsnGlnArgSerProArgSerAsnSerSerLeuAspProLys 609
Db      389 TTGGGTTTATCCAATCAAGGTCTCTCTCGCAGTAACTCATCTAGACCCCAAA 336

```

RESULT 7

```

BZ512352
LOCUS   BZ512352 520 bp DNA linear GSS 16-DEC-2002
DEFINITION BOMSL50TF BO 2.3 KB Brassica oleracea genomic clone BOMSL50, genomic survey sequence.
ACCESSION BZ512352
VERSION BZ512352.1 GI:27038385
KEYWORDS Brassica oleracea
SOURCE Brassica oleracea
ORGANISM Brassica oleracea
REFERENCE 1 (bases 1 to 520)
AUTHORS Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.
TITLE Whole genome shotgun sequencing of Brassica oleracea
JOURNAL Unpublished
COMMENT Other_GSSs: BOMSL50TR
Contact: Chris Town
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TF
Class: sheared ends.
Location/Qualifiers
1. .520
/organism="Brassica oleracea"
/mol_type="genomic DNA"
/strain="TO1000DH3"
/db_xref="taxon:3712"
/clone="BOMSL50"
/clone_lib="BO 2.3 KB"
/note="Vector: pHOSt1; Site 1: BstXI; 2-3 kb sheared genomic DNA inserted into pHOSt1 using BstXI linkers"
147 a 118 c 103 g 152 t

```

FEATURES

```

source
1. .520
/organism="Brassica oleracea"
/mol_type="genomic DNA"
/strain="TO1000DH3"
/db_xref="taxon:3712"
/clone="BOMSL50"
/clone_lib="BO 2.3 KB"
/note="Vector: pHOSt1; Site 1: BstXI; 2-3 kb sheared genomic DNA inserted into pHOSt1 using BstXI linkers"
147 a 118 c 103 g 152 t

```

BASE COUNT

ORIGIN

Alignment Scores:

RESULT 4
AV541057/c 552 bp mRNA linear EST 07-SEP-2000
LOCUS AV541057 Arabidopsis thaliana roots Columbia Arabidopsis thaliana
DEFINITION CDNA clone RZ15909F 3', mRNA sequence.
ACCESSION AV541057
VERSION AV541057.1 GI:8702815
KEYWORDS EST.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
; eurosids II; Brassicales; Brassicaceae; Arabidopsia.
REFERENCE 1 (bases 1 to 552)
AUTHORS Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
TITLE A large scale analysis of cDNA in Arabidopsis thaliana: Generation
of 12,028 non-redundant expressed sequence tags from normalized and
size-selected cDNA libraries
JOURNAL DNA Res. 7, 175-180 (2000)
MEDLINE 20363093
PUBMED 10907847
COMMENT Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
FEATURES
Location/Qualifiers
1..552
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="RZ15909F"
/issue_type="roots"
/clone_lib="Arabidopsis thaliana roots Columbia"
/note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
XhoI"
BASE COUNT 174 a 97 c 117 g 164 t
ORIGIN
Alignment Scores:
Pred. No.: 1,68e-51 Length: 552
Score: 663.00 Matches: 122
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 21.36% Indels: 0
DB: 9 Gaps: 0
US-09-890-475-1 (1-609) x AV541057 (1-552)
QY 488 AspGluileSerAlaLeuValSerTyrLeuGlyProSerThrSerPheProHisArg 507
DB 551 GATGAATATACAGCTCTTGAGTAGTACCTCGGCCGCTCAACATCTTTCTCATCGC 492
QY 508 SerArgArgSerProGlnTyrMetValProLeuProHisGlyGlyLeuGlyArgSerVal 527
DB 491 TCAGAAAGATCCCGGAATATATGGTTCCATTCACATGGTGGGTAGGAAGAAGTGA 432
QY 528 TyrAlaTyrGluHisLeuAlaProAsnSerTyrSerProGlyHisGlyHisArgLeuHis 547
DB 431 TATCATATGACATCTCGCCCAATTCATATTCCTCCAGGTCACGACATAGACTTCAT 372
QY 548 ArgGlnTyrSerProSerLeuValHisGlyGlnArgHisProLeuGlnTyrSerProPro 567
DB 371 CGACAGTACTCTCGCTCTTTGGTTCCGGACAGACATCCACTACAGTACTCTCTCCA 312
QY 568 IleHisGlyGlnGlnLeuProTyrGlyIleGlnArgValTyrArgHisSerProSer 587
DB 311 ATTCATGACACACACAGTATACCATGGTATACAAAGGTTTACAGACATTCACCATCT 252
QY 588 GluGluArgTyrLeuGlyLeuSerAsnGlnArgSerProArgSerAsnSerSerLeuAsp 607
DB 251 GAAGAAAGATATTTGGGTTTTTATCCAAATCAAAGGTCTCTCGCAGTAACATCATTAGAC 192

QY 608 ProLys 609
DB 191 CCCAAA 186
RESULT 5
BH470959 750 bp DNA linear GSS 13-DEC-2001
LOCUS BOHLJ19TR BOHL Brassica oleracea genomic clone BOHLJ19, genomic
DEFINITION survey sequence.
ACCESSION BH470959
VERSION BH470959.1 GI:17679070
KEYWORDS GSS.
SOURCE Brassica oleracea
ORGANISM Brassica oleracea
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE 1 (bases 1 to 750)
AUTHORS Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.
TITLE Whole genome shotgun sequencing of Brassica oleracea
JOURNAL Unpublished
COMMENT Other_GSSs: BOHLJ19TF
Contact: Chris Town
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TR
Class: sheared ends.
FEATURES
Location/Qualifiers
1..750
/organism="Brassica oleracea"
/mol_type="genomic DNA"
/strain="TC1000DH3"
/db_xref="taxon:3712"
/clone="BOHLJ19"
/clone_lib="BOHL"
/note="Vector: pHOS1; Site_1: BstXI; 2-3 kb sheared
genomic DNA inserted into pHOS1 using BstXI linkers"
BASE COUNT 222 a 160 c 150 g 218 t
ORIGIN
Alignment Scores:
Pred. No.: 1.43e-46 Length: 750
Score: 612.00 Matches: 135
Percent Similarity: 70.22% Conservative: 23
Best Local Similarity: 60.00% Mismatches: 34
Query Match: 19.72% Indels: 35
DB: 28 Gaps: 2
US-09-890-475-1 (1-609) x BH470959 (1-750)
QY 313 LeuValProMetValSerGlyIleValGluSerSerIleLysArgGlyMetHisIleGlu 332
DB 666 ATATTCATTCGTCGTAGGTATGTTTAACTATCAAGCTGGAAGCATATTGAA 607
QY 333 AlaLeuGluMetValTyrThrPheGlyMetGluAspLysPheSerAlaAlaLeuValLeu 352
DB 606 GCACITGGGATCATTTATACCTT-GGGATAGAGATAGTTTTCGGCTTCTTCGCTTCTA 548
QY 353 ThrSerPheLeuLysMetSerLysGluSerPheGluArgAlaLysArgLysAlaGlnSer 372
DB 547 ACTTCATTCCTAGGATGAGCAGAGATCATTTGAGAGGCAACAGAAAGCTCAAGCA 488
QY 373 ProLeuAlaPheLys----- 377
DB 487 CCGATAGCATTTGT-ATGGCCCTTCTTAGCTACCTTCATTGACTCTCTTTTTTTTCT 429
QY 378 -----GluAlaAlaThrLys 382

```

397 HisLysLeuSerProAlaLysGluLeuProGlyTrpGlnIleLysGluGlnIleValSer 416
Db CACAAAGTTAGACCCAGTGAAGAAGTACCAGGCTGGCAGATCCAGAGCAATGGCGAAG 539
417 LeuGluLysAspThrLeuGlnLeuAspLysGluMetGlu-----429
538 CTTGAGAAAGAGATTGTTACGTCGACAAACAGATGAAGAGCGAGATCCATCAGTCA 479
430 ---GluLysAlaArgSerLeuSerLeuMetGluGluAlaLysGluLysGluMetTyr 448
478 ATGGAGGAAGCGCATCCATCAGTCTAAGGAGGAGGAGCGCAATTAGCGAGAGATTGTAT 419
449 AsnGlnIleLysArgProArgLeuSerProMetGluMetProProValThrSerSer 468
418 AACCAACAGATGAACGCTCAAGGTTCTCAGAAAGGGAATGCCCAACAGCTTCCITTA 359
469 SerTyrSerProIleTyrArgAspArgSerPheProSerGlnArgAspAspGlnAsp 488
358 TCTTATCTCTATGATCCGCGACCAAGCTTCCCTAGTCACAGAGAGGAGATGCAGAT 299
489 GluLeuSerAlaLeuValSerSerTyrLeuGlyProSerThrSerPheProHisArgSer 508
298 GAAATATCAGCTCTGTGAGTAGTACCTCGGCCCATCAGCAGGTTTCTCATCGGTCA 239
509 -----ArgArgSerProGluTyrMetValProLeuProHisGlyLeuGlyArgSer 526
238 AGTCTCAGGAGATCCCTGATATATGTT-----CCACTCGTGGGTAGGAAGACT 185
527 ValTyrAlaTyrGluHisLeuAlaProAsnSerTyrSerProGlyHisGlyHisArgLeu 546
184 GTCTCTGCGTATGATCATCATCGCTCCAAATCTTATTCTCCG-----GTT 140
547 HisArgGlnTyrSerProSerLeuValHisGlyGlnArgHisProLeuGlnTyrSerPro 566
139 TCAGAGAGGTTACTTCCA-----GTTACGAGCAGAGACTTCTCAGAGTACTTCTT 86
567 ProLeHisGlyGlnGlnGlnLeuProTyrGlyLeGlnArgValTyrArgHisSerPro 586
85 CCAGTTCATGGGCAACCAACCAATGCCATATGTTCTA-----TACAGACTTCAACA 35
587 SerGluGluArgTyrLeuGlyLeuSer 595
34 TCTGAATTCACCATGGAGCTCAGC 8

RESULT 3
LOCUS BH479860
DEFINITION BOGRJ75TF BOGR Brassica oleracea genomic clone BOGRJ75, genomic survey sequence.
ACCESSION BH479860
VERSION BH479860.1 GI:17687964
KEYWORDS GSS.
SOURCE Brassica oleracea
ORGANISM Brassica oleracea
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids 1; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE 1. (bases 1 to 781)
AUTHORS Town, C.D., Van Aken, S., Uterback, T., Koo, H. and Fraser, C.M.
JOURNAL Whole genome shotgun sequencing of Brassica oleracea
COMMENT Unpublished
Other GSSs: BOGRJ75TF
Contact: Chris Town
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TF
Class: sheared ends.
Location/Qualifiers
1. 781

```

```

/organism="Brassica oleracea"
/mol_type="genomic DNA"
/strain="TO1000DH3"
/db_xref="taxon:3712"
/clone="BOGRJ75"
/clone_lib="BOGR"
/note="Vector: PHOS1; Site 1: BstXI; 2-3 kb sheared genomic DNA inserted into PHOS1 using BstXI linkers"
BASE COUNT 188 a 179 c 239 g 175 t
ORIGIN
Alignment Scores:
Pred. No.: 1,46e-53 Length: 781
Score: 687.50 Matches: 144
Percent Similarity: 70.54% Conservative: 38
Best Local Similarity: 55.81% Mismatches: 57
Query Match: 22.15% Indels: 19
DB: 28 Gaps: 3
US-09-890-475-1 (1-609) x BH479860 (1-781)
QY 28 ArgArgArgGluLeuProLysIleValGluThrGluSerThrSerMetAspIleThrIle 47
Db 19 CGAGGAGAGAGCGAGGCTACTGTGAAACCGTGCTTACAAACATCGAAACACGATC 78
QY 48 GlyGlnSerLysGlnProGlnPheLeuLysSerIleAspGluLeuAlaAlaPheSerVal 67
Db 79 GAACATCTAACACCTCAGTTTTTGAATCCATGCTGACTTAACCGGTAGCAGCC 138
QY 68 AlaValGluThrPheLysArgGlnPheAspAspLeuGlnLysHisIleGluSerIleGlu 87
Db 139 GCAGTGAACGCTTCAACCGCGCTACGACGAACTGCGAGCCACATGATGATACATCG 198
QY 88 AsnAlaIleAspSerLysLeuGluSerAsnGlyValValLeuAlaAlaArgAsnAsn 107
Db 199 AACCGATCATCTCAATCTCAAACTAACGGATCGTCAAAATCGCC-----246
QY 108 PheHisGlnProMetLeuSerProProArgAsnValSerValGluThrThrValThr 127
Db 247 -----GCCGTGTCGCTCGCGGAGAGTCTCTGGAGAAACACGCCAGCGCG 294
QY 128 Val-----SerGlnProSerGlnGluIleValProGluThrSerAsnLysProGlu 144
Db 295 ATTGCGTGCCATCGCGGCCAAACAG-----AAGTCCGAA 330
QY 145 GlyGlyArgMetCysGluLeuMetCysSerLysGlyLeuArgLysTyrIleTyrAlaAsn 164
Db 331 GCAGAGCGATTGTGCGAGTCGATGTGTAGCAAGAGCTCCGACAGGTACATGTCGTGAAC 390
QY 165 IleSerAspGlnAlaLysLeuMetGluGluIleProSerAlaLeuLysLeuAlaLysGlu 184
Db 391 ATATCTGAGAGACCAAGCTAATCGAAGAGCTCTCTGGAGCGTTGAAGCTTCCCAAGGAC 450
QY 185 ProAlaLysPheValLeuAspCysIleGlyLysPheTyrLeuGlnGlyArgAlaPhe 204
Db 451 CCGCGAAGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 510
QY 205 ThrLysGluSerProMetSerSerAlaArgGlnValSerLeuLeuIleLeuGluSerPhe 224
Db 511 GCCAACGACTCGCCCGCATCAACCGGAGAGGTTTCGCTTCTTCTTCTGAGGTGTAT 570
QY 225 LeuLeuMetProAspArgGlyLysGlyValValIleLeuSerTrpIleLysAspGlu 244
Db 571 CTCTGACGTTTGTATCTCTGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 630
QY 245 AlaGluThrAlaAlaValAlaTrpArgLysArgLeuMetThrGluGlyGlyLeuAlaAla 264
Db 631 CGCGAGGCGGCTGCTGTGTTGGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 690
QY 265 AlaGluLysMetAspAlaArgGlyLeuLeuValAlaAlaCysPheGlyVal 282
Db 691 CGCGAGGCTGTGATGCAAGGGTTTGTCTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTT 744

```

REFERENCE 1 (bases 1 to 704)
 Town, C.D., Van Aken, S., Utterback, T., Koo, H. and Fraser, C.M.
 Whole genome shotgun sequencing of *Brassica oleracea*
 TITLE
 JOURNAL
 COMMENT
 Other_GSSs: BOHBH16TF
 Contact: Chris Town
 TIGR
 9712 Medical Center Drive, Rockville, MD 20850, USA.
 Tel: 301-838-3523
 Fax: 301-838-0208
 Email: cdtown@tigr.org
 DNA is from a doubled haploid provided by Tom Osborn.
 Seq primer: TR
 Class: sheared ends.

FEATURES
 source
 1. .704
 Location/Qualifiers

/organism="Brassica oleracea"
 /mol_type="genomic DNA"
 /strain="T01000DH3"
 /db_xref="taxon:3712"
 /clone="BOHBH16"
 /note="BOHBH16"
 /note="vector: pHOS1; Site 1: BstXI; 2-3 kb sheared
 genomic DNA inserted into pHOS1 using BstXI linkers"
 BASE COUNT 173 a 165 c 157 g 209 t

Alignment Scores:
 Pred. No.: 9,72e-59 Length: 704
 Score: 742.50 Matches: 157
 Percent Similarity: 74.80% Conservative: 27
 Best Local Similarity: 63.82% Mismatches: 35
 Query Match: 23.92% Indels: 28
 DB: 28 Gaps: 7

US-09-890-475-1 (1-609) x BH491272 (1-704)

Qy 372 SerProLeuAlaPhe-----LysGluAlaAlaThrLysGlnLeuAlaVal 386
 Db 701 TCGGTGTTGGCTTTTATTCTCAACAGAAACTCGAAGCAAGAAAGCAGTTGGATGCG 642
 Qy 387 LeuSerValMetGlnCysMetGluThrHisLysLeuAspProAlaLysGluLeuPro 406
 Db 641 TTATCATCATGATGAAGTGTGGAGCTCACAAGTTAGACCACTGAAGAGTACCA 582
 Qy 407 GlyTrpGlnLysGluGlnLysValSerLeuGluLysAspThrLeuGlnLeuAspLys 426
 Db 581 GGGTGGCAGATCCAGAGCAAAATGGCCAGCTTGAGAGAGATTGTTGAGCTCGACAAA 522
 Qy 427 GluMetGlu-----GluLysAlaArgSerLeuSerLeuMet 438
 Db 521 CAGATGGAGAAAGCGAGATCCATCAGTCGAATGGAGAGCGGATCCATCAGTCTAAGG 462
 Qy 439 GluGluAlaLeuAlaLysArgMetTyrAsnGlnGlnLysArgProArgLeuSer 458
 Db 461 GAGGAAGCGCAATTAGCGAGAGATTGTATACCAAGATGAACGTCAAGTTGTCA 402
 Qy 459 ProMetGluMetProProValThrSerSerTyrSerProLeuTyrArgAspArgSer 478
 Db 401 GAAAGGGAATGCCACCAACAGCTTCCTTATCTCTATCTCTATGACCGCACCAAGC 342
 Qy 479 PheProSerGlnArgAspAspGlnAspGlnLysSerAlaLeuValSerSerTyrLeu 498
 Db 341 TTCCTATGTCACAGAGGGGAGTGCAGATGAAATATCAGCTCTTGTGAGTAGTTACCTC 282
 Qy 499 GlyProSerThrSerPheProHisArgSer-----ArgArgSerProGluTyrMetVal 516
 Db 281 GGCCCATCAGCAGGTTTTCCTCATCGTCAAGTCTCAGGAGATCCCTGAATATATGTT 222
 Qy 517 ProLeuProHisGlyGlyLeuGlyArgSerValTyrAlaTyrGluHisLeuAlaProAsn 536
 Db 221 -----CCACCTGTGGGTAGSAGAAAGTGTCTCTCGGTATGATCATCAGCTCCAAAT 168

Qy 537 SerTyrSerProGlyHisGlyHisArgLeuHisArgGlnTyrSerProSerLeuValHis 556
 Db 167 TCTTATTTCTCCG-----GTTTCAAGAGGTACTCTCCA-----GTTTCAAC 129
 Qy 557 GlyGlnArgHisProLeuGlnTyrSerProLeuHisGlyGlnGlnLeuProTyr 576
 Db 128 GGACAGAGACTTCTCCAGAGTACTCTTTCAGTTCATGGCAACCAATGCCATAT 69
 Qy 577 GlyIleGlnArgValTyrArgHisSerProSerGluGluArgTyrLeuGlyLeuSerAsn 596
 Db 68 GGTCTA-----TACAGACATTCACCATCTGT-GAAGATATCTTGGCTTTGTCCTCAAT 19
 Qy 597 GlnArgSerProArgSer 602
 Db 18 CACAGGACTCTCTGTAAC 1

RESULT 2
 BZ035667/c 706 bp DNA linear GSS 09-OCT-2002
 LOCUS ced83e10.g1 B.oleracea002 Brassica oleracea genomic, genomic survey
 DEFINITION sequence.
 ACCESSION BZ035667
 VERSION BZ035667.1 GI:23612705
 KEYWORDS GSS.
 SOURCE Brassica oleracea
 ORGANISM Brassica oleracea

REFERENCE 1 (bases 1 to 706)
 AUTHORS Delehaunty, K., Fewell, G., Fulton, L., McCombie, W.R., Miner, T., Nash
 W., Rabinowicz, P.D. and Wilson, R.K.
 TITLE Whole genome shotgun reads from *Brassica oleracea*
 JOURNAL Unpublished
 COMMENT Contact: Richard K. Wilson
 Genome Sequencing Center
 Washington University School of Medicine
 Email: submissions@watson.wustl.edu
 Plate: ced83 row: e column: 10
 Seq primer: -28RPpOT reverse
 Class: shotgun
 High quality sequence start: 17
 High quality sequence stop: 551
 Location/Qualifiers

FEATURES
 source

1. .706
 /organism="Brassica oleracea"
 /mol_type="genomic DNA"
 /db_xref="taxon:3712"
 /clone_lib="B.oleracea002"
 /notes="vector: pOTw13; Whole genome shotgun library from
 flowering buds. DNA was purified from a crude nuclear
 prep using *Brassica oleracea* T01000DH3 buds provided by
 Thomas Osborn at the University of Wisconsin. Genomic
 DNA was provided by Pablo Rabinowicz (CSHL) and the
 shotgun library prepared at Washington University Genome
 Sequencing Center."
 BASE COUNT 177 a 164 c 159 g 204 t 2 others

Alignment Scores:

Pred. No.: 2.05e-54 Length: 706
 Score: 696.00 Matches: 146
 Percent Similarity: 75.11% Conservative: 26
 Best Local Similarity: 63.76% Mismatches: 35
 Query Match: 22.42% Indels: 23
 DB: 28 Gaps: 6

US-09-890-475-1 (1-609) x BZ035667 (1-706)

Qy 377 LysGluAlaAlaThrLysGlnLeuAlaValLeuSerValMetGlnCysMetGluThr 396
 Db 657 AAACCTCGGAACGAAAGCAGTGTGCTTATCATCAGTATGAGTGTGTT-GAAGCT 599

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

3M protein - nucleic search, using frame_plus_p2n model

run on: August 25, 2003, 10:20:35 ; Search time 3509 Seconds
(without alignments)
4218.132 Million cell updates/sec

Title: US-09-890-475-1
Perfect score: 3104
Sequence: 1 MSNPPVAAQPTTANPLL.....RYLGSLNQSPRSNSLDPK 609

Scoring table:
BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Rgapop 6.0, Rgapext 7.0
Delop 6.0, Delext 7.0

Searched: 22781392 segs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame_p2n.model -DEV=xlh
-Q/cgn2_1/USPTO_SPOOL/US09890475/runat 14082003 090350 19554/app_query.fasta_1.775
-DB=EST -QFMT=fastap -SUFFIX=est -MINMATCH=0.1 -LOOFCU=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=spt -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09890475 @CGN 1 1 2810 @runat 14082003 090350 19554 -NCPU=6 -ICPU=3
-NO WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAEXT=0.5 -FGAPOP=6
-FGAEXT=7 -YGAPOP=10 -YGAEXT=0.5 -DELOP=6 -DELEXT=7

Database :
1: em estba.*
2: em esthum.*
3: em estin.*
4: em estmu.*
5: em estov.*
6: em estpl.*
7: em estro.*
8: em htc.*
9: gb est1.*
10: gb est2.*
11: gb htc.*
12: gb est3.*
13: gb est4.*
14: gb est5.*
15: em estfun.*
16: em estom.*
17: em gss hum.*
18: em gss inv.*
19: em gss pln.*
20: em gss vrt.*
21: em gss fun.*
22: em gss nam.*
23: em gss mus.*
24: em gss pro.*
25: em gss rod.*
26: em gss phg.*
27: em gss vrl.*
28: gb gss1.*

29: gb_gss2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	742.5	23.9	704	28	BH491272 BOHBH16TR
C 2	696	22.4	706	28	BZ035667 oed83610.
C 3	687.5	22.1	781	28	BH479860 BOGRJ75TF
C 4	663	21.4	552	9	AV541057 AV541057
C 5	612	19.7	750	28	BH470959 BOHLJ19TR
C 6	543	17.5	635	28	B77633 T29E23TR TA
C 7	487	15.7	520	29	BZ512352 BOMSL50TF
C 8	486.5	15.7	785	13	BQ115269 EST600845
C 9	467.5	15.1	557	10	BF050953 EST436111
C 10	440	14.2	771	12	B1176101 EST517184
C 11	435.5	14.0	688	28	BH984132 cdf42a06.
C 12	423.5	13.6	671	12	BG890481 EST516332
C 13	385	12.4	685	9	AW738032 EST339459
C 14	373	12.0	651	9	AW738612 EST340039
C 15	371	12.0	637	13	BQ118418
C 16	353.5	11.4	540	10	BE471538 EST416391
C 17	352	11.3	500	13	BQ511650 EST619065
C 18	321.5	10.4	2069	11	AY109393 Zea mays
C 19	295	9.5	544	10	BF053940 EST439170
C 20	279.5	9.0	430	10	BG097426 EST462041
C 21	277.5	8.9	765	14	CB970934 CAB10004
C 22	273	8.8	511	10	BE611371 sq80e12.y
C 23	266	8.6	421	29	CC052238 SALK_0153
C 24	257	8.3	487	29	CNS000TU3
C 25	253.5	8.2	945	14	CA933488 Arabidops
C 26	245	7.9	595	13	BQ853859 QGB21K22.
C 27	241.5	7.8	657	14	CB342143
C 28	231.5	7.5	774	10	BG48905 EST510524
C 29	231	7.4	858	10	BE661560 EST9 GmaxS
C 30	230.5	7.4	687	14	CA933310 MIUSCS.P6
C 31	229	7.4	645	14	CD480186 eca01-21m
C 32	227.5	7.3	777	10	BQ867073 QGC9L12.Y
C 33	225	7.2	750	13	BQ867073 QGC9L12.Y
C 34	222.5	7.2	594	13	BU763866 sas49b05.
C 35	218	7.0	747	29	CC344181 OGUBK77TH
C 36	217.5	7.0	800	14	CD440778 EL01N0560
C 37	216.5	7.0	810	14	CB677353 OSUNEe4D
C 38	215.5	6.9	877	12	BM779700 EST590276
C 39	214.5	6.9	723	13	BQ996736 QGG13K11.
C 40	213.5	6.9	799	12	BM779770 EST590346
C 41	213	6.9	677	10	BG593384 EST492074
C 42	211.5	6.8	652	12	BJ464949 BJ464949
C 43	211.5	6.8	678	14	CB092738
C 44	210	6.8	599	13	BU042940 PP_LFA001
C 45	208.5	6.7	782	14	CB289066 V-B-112D0

ALIGNMENTS

RESULT 1
BH491272/c
LOCUS BOHBH16TR BOHB Brassica oleracea genomic clone BOHBH16, genomic
DEFINITION survey sequence.
ACCESSION BH491272
VERSION BH491272.1 GI:17699376
KEYWORDS GSS.
SOURCE Brassica oleracea
ORGANISM Brassica oleracea
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
; eurosids II; Brassicales; Brassicaceae; Brassica.

QY 135 IleValProGluThrSerAsn----- 141
 Db 3050 CTCATGAAGACACGGCAGCTGGCCAGATCTCTCCAAACCCCAAGATCATGCCAAGGCA 3109
 QY 142 ---LysProGluGlyGlyArgMetCysGluLeuMetCysSerLysGlyLeuArgLysTyr 150
 Db 3110 CCAAGAGGTGAAGAGGCAAAATCACTAAATGCGCTGCCAGTCATTACAAACAGAACCA 3169
 QY 161 IleTyrAlaAsnIleSerAspGlnAlaLysLeuMetGluGluIleProSerAlaLeuLys 180
 Db 3170 ATAAACACCCCAACACACACAAACACAGTTG-----AAGGCATCC 3211
 QY 181 LeuAlaLysGluProAlaLysPheValLeuAspCysIleGlyLysPhe---TyrLeuGln 199
 Db 3212 CTGGGAAAGTAGGTGTGAAGAGAGCTCTAGCAGTCGGCAAGTTTCACACGAGCGTCA 3271
 QY 200 GlyArgAlaPheThrLysGluSerProMetSerSerAlaArgGlnValSerLeu 219
 Db 3272 GGGGAGACACCCACACACACAGAGCCAGCAGAGATGCCAAGATC----- 3322
 QY 220 IleLeuGluSerPheLeuLeuMetProAsp-----ArgGlyLys 232
 Db 3323 -----AGAACGTTTAAAGGAGTCTCCAAAGCAGATCTCTGGACCCAGCAGCCCGTAACT 3376
 QY 233 GlyLysValLysIleGluSerTrpIleLysAspGluAlaGluThrAlaAla---ValAla 251
 Db 3377 GGAATGAAGAGTGCCCAAGAACCGCTTAAGGAGAGGCGCCAGTCCTAGAGACCTGGCT 3436
 QY 252 TrpArgLysArgLeuMetThrGluGlyGlyLeuAlaAlaGluLysMetAspAlaArg 271
 Db 3437 GGCCTCAAGAGCTCTCCAGACACACAGTCCCTCTGAGGAATCAATGACTGATGAGAAA 3496
 QY 272 GlyLeuLeuLeuValAlaCysPheGlyValPro----- 283
 Db 3497 -----ACTACCAAAATAGCTGCAAAATCTCCACCCAGCAATCAGTGGACACTCCAACA 3550
 QY 284 -----SerAsnPheArgSerThrAspLeu-----Leu 292
 Db 3551 AGCACAAACATGGCTTAAGAGAGTCTCAGGAAGCAGATGTAGAGGAAGATTTCTTA 3610
 QY 293 AspLeuIleArgMetSerGlySerAsnGluIleAlaGlyAlaLeuLysArgSerGlnPhe 312
 Db 3611 GCACCTCAGGAAACTAACACCATCA-----GCAGGG-----AAAGCCATG 3649
 QY 313 IleValProMetValSerGlyIleValGluSerSerIleLys-----ArgGlyMetHis 330
 Db 3650 CTTACCCCAACACAGCAGAGGTGATGAGAAAGCAATTAAGCATTTATGGGAACCTCCA 3709
 QY 331 IleGluAlaLeuGluMetValTyrThrPheGlyMetGluAspLysPheSerAlaAlaLeu 350
 Db 3710 GTGCAGAAACTGGACTGGCAGAACT----- 3736
 QY 351 ValLeuThrSerPheLeuLysMetSerLysGluSerPheGluArgAlaLysArgLysAla 370
 Db 3737 -----TTACCTGGCAGCAAAACAGACAGCTACAGACTCTCTAAGGAAAGGCC 3781
 QY 371 GlnSer-----ProLeuAla---PheLysGlu----- 378
 Db 3782 CAGGCTCTAGAGAGACTGGCTGGCTTTAAAGAGCTCTTCCAGACTCTCTGGTCACACCGAG 3841
 QY 378 ----- 378
 Db 3842 GAATTAGTGGCTGCTGGTAAAAACCTAAATATACCTCGGACTCTCCACAGTCAGACCCA 3901
 QY 379 -----AlaAlaThrLysGln----- 383
 Db 3902 GTGGACACCCCAACAGCACAAGCAGCCCAAGAGAGATATCAGGAAGACAGATGTA 3961
 QY 384 -----LeuAla-----ValLeuSerSerValMetGlnCysMetGluThr 396
 Db 3962 GAGGGAGAACTCTTAGCGTGCGAGGAATCTAATGCCATCAGCGCAAGCCATGCACACG 4021
 QY 397 -----His 397

Db 4022 CCTAAACCATCAGTAGGTGAAGAGAAAGACATCATATTTTGTGGAACTCCAGTGCAG 4081
 QY 398 LysLeuAspProAlaLysGluLeuProGlyTrpGlnIleLysGluGlnIleValSerLeu 417
 Db 4082 AAATGGACCTGCAGAGAACTTAACCGGACGACAGAGACGGCCAAACTCTTAAGGAA 4141
 QY 418 GluLysAspThrLeuGlnLeuAspLysGluMetGluGluLysAlaArgSerLeuSerLeu 437
 Db 4142 GAGCCCAAGGCTCTGGAAGACCTGACTGGCTTTAAAGAGCTCTTCAGACCCCTGTGCAT 4201
 QY 438 MetGluGluAlaAlaLeuAlaLysArgMetTyrAsnGlnIleLysArgProArgLeu 457
 Db 4202 ACTGAAGAGCAGGTGGTGTGGC-----AAAACTACTAAATG 4240
 QY 458 SerProMetGluMetProProValThrSerSerSerTyrSerProIleTyrArgAspArg 477
 Db 4241 CCCTGCGAATCTTCTCCACCAATCAGACAGACCCCAACAGCAAGAGGAGCC 4300
 QY 478 SerPheProSerGlnArgAspAspGlnAspGluIleSerAlaLeuVal-SerSerTy 497
 Db 4301 AAGACACCTTTGGAGAAAGGAGCGTACAGAGAGAGCTCTCAGCCCTGAAGAAGCTCA-- 4358
 QY 497 rLeuGlyProSerThrSerPheProHisArgSerArgArgSerProGluTyrMetValPr 517
 Db 4359 -----CACACACATCAGGGGAAACCCACACACAGATATAA 4393
 QY 517 oLeuProHisGlyGlyLeuGlyArgSerValTyrAlaTyrGluHisLeuAlaProAsnSe 537
 Db 4394 -GTACCA--GGAGGTGAGGATAAAGCATCAACGGCTTTAGGGAAACTGCATAAACAGAA 4449
 QY 537 rTyrSerProGlyHisGlyHisArgLeuHisArgGlnTyrSerProSerLeuValHisG1 557
 Db 4450 ACTGAGCCCA-----GCAGCAGGTAACTGTAG 4479
 QY 557 yGlnArgHisPro 561
 Db 4480 CAAGAGGCACCCA 4492

Search completed: August 25, 2003, 15:49:25
 Job time : 3427 secs

QY 252 TrpArgLysArgLeuMetThrGluGlyGlyLeuAlaalaGluLysMetAspAlaArg 271
Db 3437 GGCTTCAAGAGCTCTTCCAGACACAGAGTCCCTCTGAGGAATCAATGACTGATGAGAA 3496
QY 272 GlyLeuLeuLeuValAlaCysPheGlyValPro----- 283
Db 3497 -----ACTACCAAAATAGCTGCAAACTCCACCACCAAGATCACTGAGACTCCAACA 3550
QY 284 -----SerAsnPheArgSerThrAspLeu-----Leu 292
Db 3551 AGCACAAGCAATGGCTTAAGAGAGTCTCAGAAAGCAGATGTAGAGGAATCTTCTTA 3610
QY 293 AspLeuIleArgMetSerGlySerAsnGluIleAlaGlyAlaLeuLysArgSerGlnPhe 312
Db 3611 GCACTCAGGAATCAACACCATCA-----GCAGGG-----AAAGCCATG 3649
QY 313 LeuValProMetValSerGlyIleValGluSerSerIleLys-----ArgGlyMetHis 330
Db 3650 CTTACGCCCAAAACAGCAGAGGTGATGAGAAAGACATTAAGCATTTATGGGAATCTCA 3709
QY 331 IleGluAlaLeuGluMetValTyThrPheGlyMetGluAspLysPheSerAlaAlaLeu 350
Db 3710 GTGCAGAACTGGACCTGGCAGGACT----- 3736
QY 351 ValLeuThrSerPheLeuLysMetSerLysGluSerPheGluArgAlaLysArgLysAla 370
Db 3737 -----TTACCTGGCAGCAAAAGACAGCTACAGACTCCTTAAGGAAAGGCC 3781
QY 371 GlnSer-----ProLeuAla-----PheLysGlu----- 378
Db 3782 CAGCTCTAGAGACTGGCTGGCTTTAAGAGCTCTTCCAGACTCTTCCAGCTCTGTCACACCGAG 3841
QY 378 ----- 378
Db 3842 GAATTAGTGGCTGCTGTAATAAACCACTAAATAACCTGGGACTCTCCACAGTCAGACCCA 3901
QY 379 -----AlaAlaThrLysGln----- 383
Db 3902 GTGCACACCCCAACAGCAGCAGCAAGCAGCAGCAAGCAAGAGTATCAGGAAGCAGATGTA 3961
QY 384 -----LeuAla-----ValLeuSerSerValMetGlnCysMetGluThr 396
Db 3962 GAGGAGAACTCTTAGCGTGCGAGGAATCTAATGCCATCAGCAGGCAAGCCATGCAACAG 4021
QY 397 -----His 397
Db 4022 CCTAACCATCAGTAGTGAAGAGAAGACATCATATTTGTGGAACTCCAGTGCAG 4081
QY 398 LysLeuAspProAlaLysGluLeuProGlyTyTrpGlnIleLysGluGlnIleValSerLeu 417
Db 4082 AAACCTGGACCTGCAGAGAACTTAACCGGCGAGCAGAGAGCGCCACAAACTCTCTAAGGAA 4141
QY 418 GluLysAspThrLeuGlnLeuAspLysGluMetGluGluLysAlaArgSerLeuSerLeu 437
Db 4142 GAGGCCAGGCTCTGGAGACCTGAGCTTTAAGAGCTCTTCCAGACCCCTGCTCAT 4201
QY 438 MetGluGluAlaAlaLeuAlaLysArgMetTyAsnGlnIleLysArgProArgLeu 457
Db 4202 ACTGAAGAAGCAGTGGCTGCTGGC-----AAACTACTAAATG 4240
QY 458 SerProMetGluMetProProValThrSerSerSerTySerProIleTyArgAspArg 477
Db 4241 CCCTGCGAATCTTCTCCACCAAGATTCAGACACACCCCAACAGCAGCAGAGGAGGCC 4300
QY 478 SerPheProSerGlnArgAspAspGlnAspGluIleSerAlaLeuVal-SerSerTy 497
Db 4301 AAGACACCTTTGGAGAAAAGGACGTCACAGAAGAGCTCTCAGCCCTGAAGAAGCTCA-- 4358
QY 497 LysGlyProSerThrSerPheProHisArgSerArgArgSerProGluTyMetValPr 517
Db 4359 -----CACAGACATCAGGGGAAACCAACACACACAGATAAA 4393

QY 517 oLeuProHisGlyGlyLeuGlyArgSerValTyAlaTyArgLysLeuAlaProAsnSe 537
Db 4394 -GTACCA--GGAGGTGAGGATAAAAGCATCAACGCGTTTAGGGAACCTGCAAAACAGAA 4449
QY 537 rTySerProGlyHisGlyHisArgLeuHisArgGlnTySerProSerLeuValHisG 557
Db 4450 ACTGGACCCA-----CGACCAAGTGTACTGTGTAG 4479
QY 557 yGlnArgHisPro 561
Db 4480 CAAGAGGCCACCCA 4492
RESULT 15
US-09-919-039-20
; Sequence 20, Application US/09919039
; Publication No. US20030108871A1
; GENERAL INFORMATION:
; APPLICANT: Kaser, Matthew R.
; TITLE OF INVENTION: GENES EXPRESSED IN TREATED HUMAN C3A LIVER CELL CULTURES
; FILE REFERENCE: PA-0035 US
; CURRENT APPLICATION NUMBER: US/09/919,039
; PRIOR FILING DATE: 2002-09-09
; PRIOR APPLICATION NUMBER: 60/222,113
; PRIOR FILING DATE: 2000-07-28
; NUMBER OF SEQ ID NOS: 401
; SOFTWARE: PERL Program
; SEQ ID NO 20
; LENGTH: 10432
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20030108871A1 2700132CB1
US-09-919-039-20

Alignment Scores: 0.0218 Length: 10432
Pred. No.: 127.50 Matches: 138
Score: 33.43% Conservative: 91
Percent Similarity: 20.15% Mismatches: 230
Best Local Similarity: 4.11% Indels: 227
Query Match: 11 Gaps: 30
DB:

US-09-890-475-1 (1-609) x US-09-919-039-20 (1-10432)

QY 11 GlnProThrThrThrAlaAsnProLeuLeuGlnArgHisGlnSerGluGlnArg--- 29
Db 2717 GAGCCTTCAAAACAGTATCCACTGTAAACAGGTCAAGGAGGTCTACAGAGTTCAGGAAT 2776
QY 30 ---ArgGluLeuProLysIleValGluThrGluSerThrSerMetAspIle----- 45
Db 2777 ATACAGAGCTACCTGTGAAAGTAAAGTGAAGAAACAAATACAGAAATTTGTGAGTGC 2836
QY 46 -----ThrIleGlyGlnSerLysGlnProGlnPheLeuLys 57
Db 2837 ATCCTAAAAGAGGTGAGGAGCAACACTACTACAAAGGAGAGAGAGAGATGAAG 2896
QY 58 SerIleAspGluLeuAlaAlaPheSerValAlaValGluThrPheLysArgGlnPheAsp 77
Db 2897 GAAATAGAA-----AGACCTTTTTCAG 2917
QY 78 AspLeuGlnLysHisIleGluSerIleGluAsnAlaIleAspSerLysLeuGluSerAsn 97
Db 2918 ACATATAAGGAAATATTTGATTAAGAAAGAAC-----GATGAAAGATGAAGCA-- 2968
QY 98 GlyValValLeuAlaAlaArgAsnAsnAsnPheHisGln-----ProMetLeuSer 114
Db 2969 -----ATGAGGAGATCAGAGACTTGGGGGCGAGAAATGTCACCAATG----- 3010
QY 115 ProProArgAsnAsnValSerValGluThrThrValThrValSerGlnProSerGlnGlu 134
Db 3011 -----TCTGACCTGCACAGACCTCAAGAGCTTCCTGATACAGAA 3049

```

QY 333 AlaLeuGluMetValTyrThrPheGlyMetGluAspLysPheSerAlaAlaLeuValLeu 352
Db 1134325 -----ACAAATGCTCGGAATACACGGGTGTGATCAACGAAACCTAT 1134284
QY 353 ThrSerPheLeu-LysMetSerLysGluSerPheGluArgAlaLysArgLysAlaGlnSe 372
Db 1134283 AGCATTGTTGTGCAAGGTTTCA-----TGCGTCTGATGGCAAGG 1134245
QY 372 rPro--LeuAlaPheLysGluAlaAlaThrLysGlnLeuAlaValLeuSerSerValMe 391
Db 1134244 CCCGATGTGACCATTTGGGAAGCATACACCGCGATCATCGTCGATCAGGAGC----- 1134190
QY 391 rClnCysMetGluThrHisLysLeuAspProAlaLysGluLeuProGlyTrpGlnLeuLy 411
Db 1134189 -AGATGGCGCAACCCACGCTCTA----- 1134166
QY 411 sGluGlnLeuValSerLeuGluLysAspThrLeuGlnLeuAspLysGluMetGluGluLy 431
Db 1134166 ----- 1134166
QY 431 sAlaArgSerLeuSerLeuMetGluGluAlaAlaLeuAlaLysArgMetTyrAsnGlnGl 451
Db 1134166 ----- 1134166
QY 451 nLeLysArgProArgLeuSerProMetGluMetProProValThrSerSerSerTyrSe 471
Db 1134165 -----CGTGGGTACCGCAACTGATGCCACCGCATGTTGCCCATTT----- 1134124
QY 471 rProLetyrArgAspArgSerPheProSerGlnArgAspAspGlnAspGluLysLe 491
Db 1134123 -----TGTTTCCGATCCGGGAC-----CT 1134101
QY 491 rAlaLeuValSerSerTyrLeuGlyProSerThrSer-----PheProHisArgSerAr 509
Db 1134100 AACGGGGTGGCCCGGAGCTATTCTTCAACGTCCCTCTGTTCCGCAT----- 1134049
QY 509 gArgSerProGluTyrMetValProLeuPro-HisGlyGlyLeuGlyArgSerValTyrA 529
Db 1134048 -----CCGGCCCATCATCGGTGGAAAGGCGG----- 1134021
QY 529 laTyrGluHisLeuAlaProAsnSerTyrSerProGlyHisGlyHisArgLeuHisArg 549
Db 1134020 -----AAACACCAAGCGGAGAGAAGTACTCTTCAACACCGCGGTGGCATGTG----- 1133973
QY 549 lnTyrSerProSerLeuValHisGlyGlnArgHisProLeuGlnTyrSerProProfile 569
Db 1133972 -----CCGAGC---GTGGAGGGCATGGGCGGGCTCAGACATCGACCT----- 1133931
QY 569 isGlyGlnGlnLeuProTyrGlyIleGlnArgValTyrArgHisSerProSerGluG 589
Db 1133930 -----CAAAGAGCTTTTCGACGCTCCCTCTCCCTC 1133900
QY 589 luArg 590
Db 1133899 AACGA 1133895

```

RESULT 14

```

US-09-919-172-97
; Sequence 97, Application US/03919172
; Patent No. US20020119463A1
; GENERAL INFORMATION:
; APPLICANT: Farris, Mary
; TITLE OF INVENTION: PROSTATE CANCER MARKERS
; FILE REFERENCE: PA-0036 US
; CURRENT APPLICATION NUMBER: US/05/919,172
; CURRENT FILING DATE: 2001-07-30
; PRIOR FILING DATE: 2000-07-28
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: PERL Program
; SEQ ID NO 97

```

```

; LENGTH: 10432
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: incyte ID No. US20020119463A1 2700132CB1
US-09-919-172-97

Alignment Scores:
Pred. No.: 0.0218 Length: 10432
Score: 127.50 Matches: 138
Percent Similarity: 33.43% Conservative: 91
Best Local Similarity: 20.15% Mismatches: 230
Query Match: 4.11% Indels: 227
DB: 10 Gaps: 30

US-09-890-475-1 (1-609) x US-09-919-172-97 (1-10432)
QY 11 GlnProThrThrAlaAsnProLeuLeuGlnArgHisGlnSerGluGlnArgArg--- 29
Db 2717 GAGCCTTCAAAACAGCATATCCACTGTAACAGGTCTAGGAGGTCTACAGAGTCAGAAAT 2776
QY 30 ---ArgGluLeuProLysIleValGluThrGluSerThrSerMetAspIle----- 45
Db 2777 ATACAGAAGCTACCTGTGGAAAGTACAGAGTGAAGAAACAAATACAGAAATTTGTGATGC 2836
QY 46 -----ThrIleGlyGlnSerLysGlnProGlnPheLeuLys 57
Db 2837 ATCTAAAGAGGTCTCAGAGGCAACACTACTACAAAGAGGAGAGAGAGATGAAG 2896
QY 58 SerIleAspGluLeuAlaAlaPheSerValAlaValGluThrPheLysArgGlnPheAsp 77
Db 2897 GAATAGNA-----AGACCTTTTGAG 2917
QY 78 AspLeuGlnLysHisIleGluSerIleGluAsnAlaIleAspSerLysLeuGluSerAsn 97
Db 2918 ACATATAAGGAAATATTTGAATTAAGAAAGAAC-----GATGAAAGATGAAGCA--- 2968
QY 98 GlyValValLeuAlaAlaArgAsnAsnAsnHisGln-----ProMetLeuSer 114
Db 2969 -----ATGAGAGATCAAGNACTTGGGGCAGAAATGTGCCACCATG----- 3010
QY 115 ProProArgAsnAsnValSerValGluThrValThrValSerGlnProSerGlnGlu 134
Db 3011 -----TCTGACCTGACAGACCTCAAGAGCTTGCTGTATACAGAA 3049
QY 135 IleValProGluThrSerAsn----- 141
Db 3050 CTCATGAAGACACGGCAGCTGGCCAGAACTCTCTCAAACCCCAAGATCATGCCAAGGCA 3109
QY 142 ---LysProGluGlyGlyArgMetCysGluLeuMetCysSerLysGlyLeuArgLysTyr 160
Db 3110 CCAAAGAGTGAGAAAGCAAAATCACTAAATGCCCTGCCAGTCATTACAAACCAAGAACCA 3169
QY 161 IleTyrAlaAsnIleSerAspGlnAlaLysLeuMetGluLulleProSerAlaLeuLys 180
Db 3170 ATAAACACCCCAACACACAAACAAACAGTTG-----AAGGCATCC 3211
QY 181 LeuAlaLysGluProAlaLysPheValLeuAspCysIleGlyLysPhe---TyrLeuGln 199
Db 3212 CTGGGAAAGTAGGTGTCAAGAGAGCTCTAGCAGTCGGCAAGTTTACACGACGCTCA 3271
QY 200 GlyArgArgAlaPheThrLysGluSerProMetSerSerAlaArgGlnValSerLeuLeu 219
Db 3272 GGGGAGACCCACGACACGACAGAGAGCAGAGATGGCAAGAGCATC----- 3322
QY 220 IleLeuGluSerPheLeuLeuMetProAsp-----ArgGlyLys 232
Db 3323 -----AGAAGTTTAAGAGTCTCCAAAGCAGATCTGGACCCAGAGCCCGGTAACT 3376
QY 233 GlyLysValLysIleGluSerTrpIleLysAspGluAlaGluThrAlaAla---Valala 251
Db 3377 GGAATGAAGAGTGGCCCAAGAAACGCTTAAGGAGAGGCCCGCCAGTCACTAGAGACCTGGCT 3436

```

Qy	496	SerTyrLeuGlyProSerThrSerPhePro-----HisArgSer 508
Db	3803	AAGGAGCCCCGGCCAAAGTGTCCATGTGCCTTCCGCCACTTACCCAGCACCAAGTCC 3862
Qy	509	ArgArgSerProGluTyrMetValProLeuProHisGlyGlyLeuGlyArgSerValTyr 528
Db	3863	-----TACATCCCTCATCATCAGCGC-----TAT 3886
Qy	529	AlaTyrGluHisLeuAlaProAsnSerTyrSerProGlyHis---GlyHisArgLeuHis 547
Db	3887	TCCTACAGT-----CAGTCTCATGACCCCAACCAACCCAGCTACCGG----- 3928
Qy	548	ArgGlnTyrSerProSerLeuValHisGlyGlnArgHisProLeuGlnTyrSerProPro 567
Db	3929	-----AGCATGCCCTGCTGTGATCATGCAGAACTACCCAGGTTCTACTCGCTTCC 3979
Qy	568	IleHisGlyGlnGlnLeuProTyrGlyIleGln----- 579
Db	3980	AGTACTCTTTTTC-----CCATATGGCAGCAAGGTCTCAGGTGGTGAAGATGCTGAC 4033
Qy	580	-----ArgValTyrArgHisSerProSerGluGluArgTyrLeuGly 593
Db	4034	AAGGCACGAGCCAGCCCGGTCGACTTGTAAATCCAGGTCAGAGTCCAAAGCCCTGGAC 4093
Qy	594	LeuSerAsnGlnArgSerProArgSerAsnSer 604
Db	4094	ATCTTGCACGACGATGCCAGTCACTACAAGAGC 4126
RESULT 13		
US-09-738-626-1/c		
; Sequence 1, Application US/09738626		
; Publication No. US20020197605A1		
; GENERAL INFORMATION:		
; APPLICANT: NAKAGAWA, SATOSHI		
; APPLICANT: MIZOGUCHI, HIROSHI		
; APPLICANT: ANDO, SEIKO		
; APPLICANT: HAYASHI, MIKIRO		
; APPLICANT: OCHIAI, KEIKO		
; APPLICANT: YOKOI, HARUHIKO		
; APPLICANT: TATEISHI, NAKO		
; APPLICANT: SENOH, AKIHIRO		
; APPLICANT: IKEDA, MASATO		
; APPLICANT: OZAKI, AKIO		
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES		
; FILE REFERENCE: 249-125		
; CURRENT APPLICATION NUMBER: US/09/738,626		
; CURRENT FILING DATE: 2000-12-19		
; PRIOR APPLICATION NUMBER: JP 99/377484		
; PRIOR FILING DATE: 1999-12-16		
; PRIOR APPLICATION NUMBER: JP 00/159162		
; PRIOR FILING DATE: 2000-04-07		
; PRIOR APPLICATION NUMBER: JP 00/280988		
; PRIOR FILING DATE: 2000-08-03		
; NUMBER OF SEQ ID NOS: 7059		
; SOFTWARE: PatentIn ver. 3.0		
; SEQ ID NO 1		
; LENGTH: 3309400		
; TYPE: DNA		
; ORGANISM: Corynebacterium glutamicum		
US-09-738-626-1		

Alignment Scores:			
Pred. No.:	108	Length:	3309400
Score:	129.00	Matches:	132
Percent Similarity:	32.50%	Conservative:	77
Best Local Similarity:	20.5%	Mismatches:	184
Query Match:	4.16%	Indels:	250
DB:	10	Gaps:	28

US-09-890-475-1 (1-609) x US-09-738-626-1 (1-3309400)

6 ProThrValAlaAlaGlnProThrThrAlaAsnProLeuGlnArgHisGlnSer 25

[illegible]

1134373 GTGTTCA CGG GTGTGT CCGGATCT GGC CAATCCTCGTGGTGTTCGGC----- 1134326

```

2000-05-17; 2000-05-17
NUMBER OF SEQ ID NOS: 159
SOFTWARE: PERL Program
SEQ ID NO 31
LENGTH: 9549
TYPE: DNA
ORGANISM: Homo sapiens
NAME/KEY: misc feature
OTHER INFORMATION: Incyte ID No. US20030124569A1 LI:350243.2:2000MAY01
FEATURE:
NAME/KEY: unsure
LOCATION: 878, 1210, 8010, 8017-8018, 8030, 8032
OTHER INFORMATION: a, t, c, g, or other
JS-10-204-887-31

Alignment Scores:
Pred. No.: 0.013 Length: 9549
Score: 129.00 Matches: 139
Percent Similarity: 30.92% Conservative: 87
Best Local Similarity: 19.02% Mismatches: 289
Query Match: 4.16% Indels: 216
JB: 14 Gaps: 27

JS-09-890-475-1 (1-609) x US-10-204-887-31 (1-9549)
2Y 2 SerAsnTyrProThrValAlaAlaGlnProThrThrThrAlaAsnProLeuLeuGln 21
Db TCCAGACAGCCACCTTCACTAGCAGCAGCCAGCCAGGCTTCTCCAGGCTTGACGGCCACA 2257
2Y 22 ArgHisGlnSerGluGlnArgArgGluLeuProLysIleValGluThr----- 38
Db GTGGCACAAGCCATCCCTCAGCTCCCACTCAAGCCCATTCAGCCCATAGCCCACTGT 2317
3Y -----GluSerThrSerMetAspIleThrIleGlnSerLysGln 52
Db TATGGGAGAACCTTTACAGTCAACCTGCCTTGACTCCAGCCAGGACCAAGAAAGAA 2377
3Y 53 ProGlnPheLeuLysSerIle-----AspGlu 61
Db AGACAAATAAGAGGAATCTTCAAGGAATCTTAAGTCTCTGACCCCTGGGACACC 2437
3Y 62 LeuAlaAlaPheSerValAlaValGluThrPheLysArgGlnPheAspLeuGlnLys 81
Db CTTGTGCGAGCAGATGTAAGTCAATTTATAGCCCATTCAGGGAATCTTCAGGAATGGGATG 2497
3Y 82 HisIleGluSerIleGluAsnAlaIleAspSerLysLeuGluSerAsnGlyValValLeu 101
Db AAAATGGAGGGCTCTCTAAATGGCTCATCACACCCCAACCAAGCCGACTGGCTAGCATC 2557
3Y 102 AlaAlaArgAsnAsnAsnPheHisGlnProMetLeuSerProProArgAsnValSer 121
Db AAGGCTGAAGCCGACAAGATCTACAGTTTCAAGGACAATGCCCCAGCCCTTCCATTGGA 2617
3Y 122 ValGluThrThrValThrValSerGlnProSerGlnIleValPro-----Glu 138
Db GGCAGTAGCCGCTTGAAGAACATCCCTTACTCAGCCCTTGACTCTCTTACATGTGGTG 2677
3Y 139 ThrSerAsnLysProGluGlyGlyArgMetCysGluLeuMetCysSerLysGlyLeuArg 158
Db ACCAGATGGAGCTGAAGCCAGC-----TCAGTCAAAAACCAAC 2716
3Y 159 LysTyrIleTyrAlaAsnIleSerAspGlnAlaLys-----LeuMetGluGlu 174
Db AGCCCTGCATCTCTGACATCTCTGATCTGGGAGGATGGGGAGGCAAGGTAGACAGT 2776
3Y 175 IleProSer-----AlaLeuLysLeuAlaLysGluProAlaLysPheValLeu----- 190
Db GTCAATCAAGGACGCCGACAGTGTGTAAAGAGGGGCTTAAGAAAATCTTTTCC 2836
3Y 191 -----AspCysIleGlyLysPheTyrLeuGlnGlyArgAlaPheThrLysGlu 207
Db CCTCAGCCTCAGAGCAAGACTCACCATATTACCAGGCTTTGAGAGTTACTATTCTCCA 2896
208 Ser-----PrometSerSerAlaArgGlnValSer 217
Db AGTTATGCACAGTCCAGCCCTGGGGCTCTGAACCCAGCAGCAGCAGCAGCA----- 2947
218 LeuLeuIleLeuGluSerPheLeuLeuMetProAspArgGlyLysGlyLysValLysIle 237
Db -----GTGGAGAGCCAGGCCCTGAAGACAAAGGATGAGAACCTGAGAGCATA 2998
2948 -----
238 GluSerTrpIleLysAspGluAlaGluThrAlaAlaValAlaTrpArgLysArgLeuMet 257
Db -----
2999 GAAGGGAAGTGAAGAACCAT----- 3019
258 ThrGluGlyGlyLeuAlaAlaAlaGluLysMetAspAlaArgGlyLeuLeuLeuVal 277
Db ----- 3019
278 AlaCysPheGlyValProSerAsnPheArgSerThrAspLeuLeuAspLeuIleArgMet 297
Db ATCTGTGAAGAAAGAGCCGAGCTGAGCAGTTCCAGTCAGCAGCCCTCGCTCATCCAG 3079
298 SerGlySerAsnGluIleAlaGlyAlaLeuLysArgSerGlnPheLeuValProMetVal 317
Db CAGCGTCCCAATATGTACATGCGAGTCCCTGTACTACACCATGCTATGTATCCCCC 3139
318 SerGlyIleValGluSerSerIleLysArgGlyMetHisIleGluAlaLeuGluMetVal 337
Db TATGGCTACAGCAGCAGCAGCTTACCAC-----ACCCACCTTCTGAGCACAACACGGCT 3193
338 TyrThrPheGlyMetGluAspLysPheSerAlaAlaLeuValLeuThrSerPheLeuLys 357
Db TACCGCAGCAGCAGTACGAGAA-----CAG 3217
358 MetSerLysGluSerPheGluArgAlaLysArg-----LysAlaGlnSerPro 373
Db CAGAAACGCCAGAGCTTAGAGCAGCAGCAGCGGGAGTGCGCAAGAGCAGAGATGGGC 3277
374 LeuAlaPheLysGluAlaAlaThrLysGln----- 383
Db CTGAAGGAGCGGGAGGAGCAGCTCAAGGAAGAGTGGAAAGCCGCAATTCACCA 3337
394 --LeuAlaValLeuSerSerValMetGlnCysMetGluThr----- 396
Db ACTCTCACCAGGCCCCAGCCTGACAGCAGCTGTGAATCAGGACCTGGCAAGGCCAAG 3397
397 -----HisLysLeuAspProAlaLysGlu 404
Db GAGCAGGGGCTGACCAGCCAAATCAGTCATCATCTCCCAAGTTAGTACTCTTCANAA 3457
405 LeuProGlyTrpGlnIleLysGluGlnIleValSerLeuGluLysAspThrLeuGlnLeu 424
Db CTCCCGGC-----CAGGCCCTGAAGCCTTAAAGTGAAGCTGAGTATGCCAGCCACTA 3514
425 AspLysGluMetGluGluLysAlaArgSerLeuSerLeuMetGluGluAlaAlaLeuAla 444
Db AGCAAGGAGGCTCTGAGGCCAAGACAGAGTGCTGAGTGTGTCGACAGGCGAGAGATGGAT 3574
445 LysArgMetTyrAsnGlnGlnIleLysArgProArgLeuSerProMetGluMetProPro 464
Db CCAATATCTCTGGTACCAGCAGAGGCGAGCCCGGATGCGGACATATGTTTATCTT----- 3631
465 ValThrSerSerTyrSerProIle-----TyrArg 475
Db -----GCCAAGTACTCAGACATCAAGTCAGAGGATGAGCGGTGGAGGAGGAGCGG 3682
476 AspArgSerPheProSerGlnArgAspAspGlnAspGluIle----- 490
Db GACCGCAATTAAGGAGGAAGGAGTCGAGTAAGGACTCTGTCCCCCAAGGAAGATGGG 3742
491 -----SerAlaLeuValSer 495
Db AAGGAAGCACAAAGTAGTACTGCAAGCTGCCACGTCAGAGGAGTCTCCGCTTGGGAGC 3802

```



```
QY 424 LeuAspLysGluMetGluLysAlaArgSerLeuMetGluLysAlaLeu 443
Db 2130 TACACTCAAGTTCGAGAGGCTACCTCTTACTAGCCC-CGTCTCGAAGGCGCCTT 2188
QY 444 AlaLysArgMetTyrAsnGlnGlnLeuLysArgProArgLeuSerProMetGluMetPro 463
Db 2189 CCCAAGAGAGCTATGC-AGCTCCCTGAGGAGACCA--TTGCTGCCATCGACGTGCTCT 2244
QY 464 ProValThrSerSerTyrSerProLysTyrArgAspArgSerPheProSerGln-Ar 483
Db 2245 TCGAGAACATCCGAGATTCAGCGCGCCGACGAGAGAGAGC---CCCTCCAGGTGC 2301
QY 483 gAspAspGlnAspGluLysSerAlaLeuValSerSerTyrLeuGlyProSerThrSe 503
Db 2302 AGACCATGCCCGGTGTGTCTGCAGCGCTTCTCTCGTCCCATCGAGCGCGTGCCT 2361
QY 503 rPheProHisArgSerArgArgSerProGluTyrMetValProLeuProHisGlyGlyLe 523
Db 2362 ACATCCCGCGGTACCGCGCTTCTCCCGAGCAGC-----TGCCCTTATGCTGGGTG 2412
QY 523 uGlyArgSerValTyrAlaTyrGluHisLeuAlaProAsnSerTyrSerProGlyHisGl 543
Db 2413 TTCCCGC-----CATGG 2424
QY 543 yHisArgLeuHisArgGlnTyrSerProSerLeuValHisGlyGlnArgHisProLeuGl 563
Db 2425 TCGCCGCGCTG-----CAACAAGATTGTGTCTG 2451
QY 563 nTyrSerProPro-----IleHisGlyGlnGlnLeuPro----- 575
Db 2452 CCTCTCTCCCGCGCGGAGCAACCATCCTCCGAGATTGTCACGCTGCCTCACAGG 2511
QY 576 -TyrGly--IleGlnArgValTyrArgHisSerProSer----- 587
Db 2512 TTGGCGCGAGTCCATCGTCTTGGCGCGGTGCGCCAGCCCTAGCTGCCTGCTACG 2571
QY 588 ----GluGluArgTyrLeuGlyLeuSerAsnGlnArgSerProArgSerAsnSerSerLe 606
Db 2572 GCACCGAGAGCATCACCAAGTCGACAGATTCTCGGCC-CCGGTAACCAAGTTCGTCTACT 2630
QY 606 uAppProLys 609
Db 2631 GCTGCCAAGA 2640
```

RESULT 10

US-10-037-270-372

Sequence 372, Application US/10037270

Publication No. US20030104529A1

GENERAL INFORMATION:

APPLICANT: Tang, Y. Tom

APPLICANT: Liu, Chenghua

APPLICANT: Asundi, Vinod

APPLICANT: Zhang, Jie

APPLICANT: Ren, Feiyan

APPLICANT: Chen, Rui-hong

APPLICANT: Zhao, Qing A.

APPLICANT: Wehrman, Tom

APPLICANT: Xue, Aidong J.

APPLICANT: Yang, Yonghong

APPLICANT: Wang, Jian-Rui

APPLICANT: Zhou, Ping

APPLICANT: Ma, Yunqing

APPLICANT: Wang, Dunrui

APPLICANT: Wang, Zhiwei

APPLICANT: Tillinghast, John

APPLICANT: Drmanac, Radoje T.

TITLE OF INVENTION: No. US20030104529A1el Nucleic Acids and

FILE REFERENCE: 784CIP2B

CURRENT APPLICATION NUMBER: US/10/037,270

CURRENT FILING DATE: 2002-01-04

PRIOR APPLICATION NUMBER: 09/552,317

PRIOR FILING DATE: 2000-04-25

```
/ PRIOR APPLICATION NUMBER: 09/488,725
/ PRIOR FILING DATE: 2000-01-21
/ NUMBER OF SEQ ID NOS: 1104
/ SOFTWARE: pt_FL_genes Version 1.0
/ SEQ ID NO 372
/ LENGTH: 7393
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (118)..(3012)
/ US-10-037-270-372
```

```
Alignment Scores:
Pred. No.: 0.00604 Length: 7393
Score: 130.50 Matches: 137
Percent Similarity: 30.80% Conservative: 86
Best Local Similarity: 18.92% Mismatches: 274
Query Match: 4.20% Indels: 227
DB: 14 Gaps: 29
```

US-09-890-475-1 (1-609) x US-10-037-270-372 (1-7393)

```
QY 1 MetSerAsnTyrPro-----ProThrValAlaAlaGlnPro 12
Db 859 ATGCCCAACAGTCCCAACTCAAGCCCATTCAGCCCAAGCCCACTGTTATGGGAGAACCT 918
QY 13 ThrThrThrAlaAsnProLeuLeu-----GlnArgHisGlnSerGluGlnArg 28
Db 919 ----TTCAACAGTCAACCTGCTTGACTCCAGCCAGGACAGAAAGAAAGACACAAAA 975
QY 29 ArgArgGluLeuProLysLeuValGluThrGluSerThrSerMetAspIleThrIleGly 48
Db 976 AAGAAGGAATCTTCAAGAGAACTTGAAGTCTCTGACCCCTCGGAAAGGTGTGTGAGCA 1035
QY 49 GlnSerLysGlnProGlnPheLeuLysSerIleAspGluLeuAlaAlaPheSerValAla 68
Db 1036 GAGGAAGCGCAAGGCCCATTC----- 1056
QY 69 ValGluThrPheLysArgGlnPheAspAspLeuGlnLysHisIleGluSerIleGluAsn 88
Db 1057 -----AGGGAATCTTCAGGAATGGGATGAAATGGAGGGGCTCTCTAAAT 1101
QY 89 AlaIleAspSerLysLeuGluSerAsnGlyValValLeuAlaAlaArgAsnAsnAsnPhe 108
Db 1102 GGCTCATAGACCCCAACAAAGCGCTGCTGCTAGCTCAAGGCTGAAGCCGACCAAGATC 1161
QY 109 HisGlnProMetLeuSerProProArgAsnAsnValSerValGluThrThrValThrVal 128
Db 1162 TACAGTTTCACGACAAATGCCCCAGCCCTTCCATGGAGGCAGTAGCGGCTTGAAGAAC 1221
QY 129 SerGlnProSerGlnGluLeuValPro-----GluThrSerAsnLysProGluGly 145
Db 1222 ACTACCCCTACTCAGCCCTGACTCCTTACATGTGTGGTGCACCCAGAAATGGAGCTGAAGCC 1281
QY 146 GlyArgMetCysGluLeuMetCysSerLysGlyLeuArgLysTyrIleTyrAlaAsnIle 165
Db 1282 AGC-----TCAGTCAAAACCAACAGCCCTGCGCATCTCTGACATC 1320
QY 166 SerAspGlnAlaLys-----LeuMetGluGluIleProSer-----AlaLeu 179
Db 1321 TCTGATGCTGGGAGGATGGGAGGCGCAAGGTAGACAGTGTCAAATCAAGAGCGCGGAA 1380
QY 180 LysLeuAlaLysGluProAlaLysPheValLeu-----AspCysIleGly 194
Db 1381 CAGTTGTTAAAGAGGGGCTAAGAAACTCTTTTCCCTCAGCCTCAGAGCAAGAC 1440
QY 195 LysPheTyrIeuGlnGlyArgAlaPheThrLysGluSer----- 208
Db 1441 TCACCATATTACCAAGGCTTTGAGAGTTACTATTCTCCAAGTTATGCACAGTCCAGCCCT 1500
QY 209 -----ProMetSerAlaArgGlnValSerLeuLeuIleLeuGluSerPhe 224
```

2048 GAGCTGCTCCGGACACAGGAGGAGCTTAAGAACTGCAGCGCAGAACGGCAGAGCCAGGAG 2107
 470 TyrSerProIleTyrArgAspArgSerPheProSerGln-----Arg 483
 2108 GTGCTCGGGCAGACACCGGACCGGAGTTGGAGAACGACTGGCGGTCTCTGAGGTTCGAG 2167
 484 AspAspAspGlnAspGluIleSerAla-LeuValSerSerTyrLeuGlyProSerThrSe 503
 2168 GCTGATCAGGTCGGGAGCTGGAAGAACACAACTCCAGCTACAAAGACCCCTCCAGCAA 2227
 503 rPheProHisArgSerArgArgSerPro 512
 2228 CTGGACAGGAGCTGTGAAGAGGCTTCCA 2255

RESULT 9

JS-09-751-962-1
 : Sequence 1, Application US/09751962
 : Patent No. US20020055165A1
 : GENERAL INFORMATION:
 : APPLICANT: CATCHESIDE, DAVID E.
 : TITLE OF INVENTION: REAGENTS AND METHODS FOR DIVERSIFICATION
 : OF DNA

NUMBER OF SEQUENCES: 2
 CORRESPONDENCE ADDRESSES:
 ADDRESSEE: Merchant, Gould, Smith, Edell, Welter & Schmidt
 STREET: 3100 No. US20020055165A1west Center, 90 South 7th Street
 CITY: Minneapolis
 STATE: MN
 COUNTRY: USA
 ZIP: 55402

COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: DOS
 SOFTWARE: FastSeq for Windows Version 2.0
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/751,962
 FILING DATE: 29-Dec-2000
 CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/977,171
 FILING DATE: <Unknown>
 ATTORNEY/AGENT INFORMATION:
 NAME: Skoog, Mark T
 REGISTRATION NUMBER: 40,178
 REFERENCE/DOCKET NUMBER: 10552.13US01
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 612-332-5300
 TELEFAX: 612-332-9081
 TELEX: <Unknown>

INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 9775 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: Genomic DNA
 SEQUENCE DESCRIPTION: SEQ ID NO: 1:
 US-09-751-962-1

Alignment Scores:
 Pred. No.: 0.00821 Length: 9775
 Score: 131.00 Matches: 128
 Percent Similarity: 33.73% Conservative: 69
 Best Local Similarity: 21.92% Mismatches: 189
 Query Match: 4.22% Indels: 203
 DB: 9 Gaps: 23

US-09-890-475-1 (1-609) x US-09-751-962-1 (1-9775)

QY 174 GluIleProSerAlaLeuLysLeuAlaLysGluProAlaLysPheValLeuAspCysIle 193
 DB 1046 GAGCGGCTTTTGTCTCTCCGGCTTCGATCA-----GACTGCTTC 1084

QY 194 GlyLysPheTyrLeuGlnGlyArgArgAlaPheThrLysGluSerProMetSerSerAla 213
 DB 1085 CG-AGCTGCACAGTTCTTGGAGAGGCGCAGAGACAAGAAATACCCCTCTCTCATC- 1142
 QY 214 ArgGlnValSerLeuLeuLeuLeuGluSerPheLeu----- 225
 DB 1143 AAGCCGCTTCTCGGGCCGATCTCGAACAGTTTCATCCAGGTGCGCGCCGCAAGGCTAACGCC 1202
 QY 226 -----LeuMetProAspArgGly-----LysGlyLysValLys 236
 DB 1203 ATCCCATCTCTGCCATCCACTGGCTTGACAAAGAGGAGGAGCGCGAAAGGTTCGCC 1262
 QY 237 IleGluSer-----Trp----- 240
 DB 1263 ATCTCCACCATCTCTCGAGCGCTCTGGAAGTCTGACCGCTCCGATGCTGTGCTCCCCACC 1322
 QY 241 ---IleLysAspGluAlaGluThrAla-----AlaValAlaTrp 252
 DB 1323 GTTGTGCTGTGATGAGCAGCAGCAGCTGCTCTGGGTCTGGTCTACAGCAGTCCCGAGAGTGTG 1382
 QY 253 ArgLysArgLeuMetThrGluGlyGlyLeuAlaAlaGluLysMetAspAlaArgGly 272
 DB 1383 AAGAGGCGCTTCAGGACACAGACTGGTGTCTATCAGAGCGCGAAG-----CGCGGT 1433
 QY 273 LeuLeuLeuValAlaLysCysPheGlyValProSerAsnPheArgSerThrAspLeuLeu 292
 DB 1434 CTCTGTGTACAAGGTGCTACTTCCGGA-----GACACTCAG 1469
 QY 293 AspLeuIleArgMetSerGlySer----- 300
 DB 1470 GAGCTGCTCGCATCTCGTTGACTCGGATAACGATGCTCTCAAGTTTGTCTGAAGCAG 1529
 QY 301 -----AsnGluIleAlaGlyAlaLeu 307
 DB 1530 AAGGTCGTTTCTGCCACCTCGATCAGTCCGCTGCTTGTGTCATCAAGGCTTCCC 1589
 QY 308 LysArgSerGlnPheLeuValProMetValSerGlyIleValGluSerSer----- 324
 DB 1590 AGCTCGAGCAGACTTTGATTTTCGAGGAACAGTCTGCCCGCGAGGGCTCTTACTACTGCC 1649
 QY 325 -----IleLysArgGlyMetHisIleGluAlaLeuMetVal 337
 DB 1650 CGTCTCTTCTCCGATGAGAAGTAGTCCGGGCCAAGATCATGAGGAGGCTGAGAGGCTC 1709
 QY 338 TyrThr-----PheGlyMetGluAspLysPheSerAlaAla 349
 DB 1710 TGCACCGCTCAGACCCCGCAGGAAATCGCCTTTGAGGCTGCCGATCTCTTCTACTTTGCT 1769
 QY 350 LeuValLeuThrSerPheLeuLysMetSerLysGluSerPheGluArgAla----- 366
 DB 1770 CTTACAGGCGCGTTGCTGCGCGGCTTACTCTTCCGATATCGAAAGGAGCCTTGACGCC 1829
 QY 366 ----- 366
 DB 1830 AAGAGCTGGAAGTCAAGCGCAGGACTGGAGATGCTAAGGTAAGTGGGCTGAGAGAGAG 1889
 QY 367 -----LysArgLysAlaGlnSerProLeuAlaPheLysGluAlaAlaThrLysGlnLeu 384
 DB 1890 GGATCAAGCTCGCGGCTCGCTCCGCTCCGCTGCCACTTCGCGCCCTGTCCACCAAGGAGGCC 1949
 QY 385 AlaValLeuSerSerValMetGlnCys-----MetGlu 395
 DB 1950 GCCCAGGAGACCCCTCGAAGATCACCATGAGAGCTTTCGACGCTCCCAAGGTCTCT 2009
 QY 396 ThrHisLysLeuAspProAlaLysGluLeuProGlyTyrGlnIleLysGluIleVal 415
 DB 2010 ACCGAGAGTCTGATGCTCTCAAGCGTCTCGCAAAAGTCGTCGATGCCATCTAC 2069
 QY 416 SerLeu-----GluLysAspThrLeuGln 423
 DB 2070 AAGATCATTTGTCCCATCATCGAGGAGCTCCGCAAGACGCGCAAGGCTGTCTGTCTCG 2129

APPLICANT: Peter VEIBY
APPLICANT: Gordon B. MILLS
APPLICANT: Robert C. EAST, Jr.
APPLICANT: Karen LU
APPLICANT: Rosemarie SCHWANDT
APPLICANT: Xumei ZHAO
APPLICANT: Karen GHATT

TITLE OF INVENTION: Nucleic Acid Molecules and Proteins For The Identification, Assessment, Prevention, and Therapy of Ovarian Cancer

FILE REFERENCE: MRI-030

CURRENT APPLICATION NUMBER: US/10/097,340

PRIOR FILING DATE: 2002-03-14

PRIOR APPLICATION NUMBER: 60/276,025

PRIOR FILING DATE: 2001-03-14

PRIOR APPLICATION NUMBER: 60/325,149

PRIOR FILING DATE: 2001-09-26

PRIOR APPLICATION NUMBER: 60/276,026

PRIOR FILING DATE: 2001-03-14

PRIOR APPLICATION NUMBER: 60/324,967

PRIOR FILING DATE: 2001/09/26

PRIOR APPLICATION NUMBER: 60/311,732

PRIOR FILING DATE: 2001-08-10

PRIOR APPLICATION NUMBER: 60/325,102

PRIOR FILING DATE: 2001-09-26

PRIOR APPLICATION NUMBER: 60/323,580

PRIOR FILING DATE: 2001-09-19

NUMBER OF SEQ ID NOS: 363

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 42

LENGTH: 5142

TYPE: DNA

ORGANISM: Homo sapiens

US-10-097-340-42

Alignment Scores:
Pred. No.: 0.00209 Length: 5142
Score: 132.50 Matches: 122
Percent Similarity: 37.37% Conservative: 91
Best Local Similarity: 21.40% Mismatches: 236
Query Match: 4.27% Indels: 122
DB: 14 Gaps: 25

US-09-890-475-1 (1-609) x US-10-097-340-42 (1-5142)

Qy 2 SerAsnTyrProThrValAlaAlaGlnProThrThrThr----- 15
Db 736 TCGAGTGGCGCGGACACGATGCTACCCCTGAAACGCGCAACGCGAGAGCGCT 795
Qy 16 AlaAsnProLeuGlnArgHisGlnSerGlnArgArgGluLeuProLysIle 35
Db 796 GGACAGCGCG-CTCCACGCGGACACCTTTGAGGAACGCGGCGCGCTCCACCAAC--- 850
Qy 36 ValGluThrGluSerThrSerMetAspIleThrIleGlyGlnSerIleGlnProGlnPhe 55
Db 851 CATGGACCTTACGACAAATAATGACACCATGTTGGGCACTTCGAAACGCGCGCGCG 910
Qy 56 LeuLysSerIleAspGluLeuAlaAlaPheSerValAlaValGluThrPheLysArgGln 75
Db 911 AGCCAGAACCTGAGTCTCTCAGTGGCTTTAGCGGTCT-----CGTCAG 955
Qy 76 PheAspAsp-----LeuGlnIysHisIleGluSerIleGluAsnAlaIleAsp---Ser 92
Db 956 ACTCAGACCTGGGTCTCTTACAGACTTTTGGAGGCGCGGAGGATGACAGACCCAC 1015
Qy 93 LysLeuGluSerAsnGlyValValLeuAlaAlaArgAsnAsnAsnPheHisGlnProMet 112
Db 1016 ATGCTGCAGTTCAAATCAACTCCAGACCTCTTCGAGAC-----CAGCAGGAG 1063
Qy 113 LeuSerProProArgAsnAsnValSerValGluThrThrVal-----ThrValSerGln 130
Db 1064 GCAGCCCCCAGGAGTGGACCATATGAAAGGCCACCATCTATGGCATCTCTAGGGAG 1123
Qy 131 ProSerGlnGluIleValProGluThrSerAsnLysProGluGlyGlyArgMetCysGlu 150

Db 1124 GGAAGCTCAGAA-----AGTGAACCTCTGTGAGGAGGAGGTTAGTTTGGTCTGGAG 1177
Qy 151 -----LeuMetCysSerLysGlyLeuArgLysTyrIleTyrAla----- 163
Db 1178 AAGATGAGCCTCTAGTGATGGTTTCTTCTGGTTCTACTAAGCCCTGCGCAGGCGAGGCT 1237
Qy 164 -----AsnIleSerAsp---GlnAlaLysLeuMetGluGluIle----- 175
Db 1238 GAGCTTACCAGAAAGTGGAGGAGCTACAGCGAAAGCTGGATGAAGAGGTGAAGAAGCGG 1297
Qy 176 -----ProSerAlaLeuAlaLysLeuAlaLysGluProAlaLysPheValLeuAsp 191
Db 1298 CAGAAGCTAGAGCCATCCCAAGTTGGCTGGAGCGGCGAGCTGGAGAGAAACAGAGAG 1357
Qy 192 CysIleGlyLysPheTyrLeuGlnGlyArgAlaPheThrLysGluSerProMetSer 211
Db 1358 TGCAGCCGACTGCAGGAGCTGCTGGAGAGG-----AAGGGGAGGCCAGCAG 1408
Qy 212 SerAlaArgGlnValSerLeuLeuIleLeuGluSerPheLeuLeuMetProAspArgGly 231
Db 1409 AGCAACAAAGGAG-----CTCCAGAACATGAAGCGCTCTTGGACCGAGGT 1453
Qy 232 LysGlyLysValLysIleGluSerTrpIleLysAspGluAlaGluThrAlaAlaValAla 251
Db 1454 GAGAT-----TTACGACATGGGCTGGAGAGCCAGCGTGTGAG 1492
Qy 252 TrpArgLysArgLeuMetThrGluGlyGlyLeuAlaAlaLysLysMetAspAlaArg 271
Db 1493 CTGCAGAACAAAGCTGAAACATGTCCAGGCTCTGAGCCTCTAAGGAGGTGTACTGAAG 1552
Qy 272 GlyLeuLeuLeuValAlaCysPheGlyValProSerAsnPheArgSerThrAspLeu 291
Db 1553 GACCTGTTA-----GAGACCCGCGAACTT 1576
Qy 292 LeuAspLeuIleArgMetSerGlySerAsnGluIleAlaGlyAlaLysLys---ArgSer 310
Db 1577 CTGGAAGAGGTC---TTGGAGGGGAAACAGCGAGTAGAGGAGCGCTGCGCGAG 1633
Qy 311 GlnPheLeuValProMetValSerGlyIleValGluSerSerIleLysArgGlyMetHis 330
Db 1634 CGGAGTTGACAGCCCTGAAGGGGCTCGAAGAGGAGGAGTACCTCCCTGACCAGGAG 1693
Qy 331 IleGluAlaLeuGluMetValTyrThrPheGlyMetGluAspLysPheSerAlaAlaLeu 350
Db 1694 GTGGAAACATGTCCGCGCAGCAGTAC----- 1717
Qy 351 ValLeuThrSerPheLeuLysMetSerLysGluSerPheGluArgAlaLysArgLysAla 370
Db 1718 -----CAGCGACACAGAGCAGCTCCGAGGAGCATGCAAGATGCA 1759
Qy 371 GlnSerProLeuAlaPheLysGluAlaThrLysGlnLeuAlaValLeuSerSerVal 390
Db 1760 ACCCAGGACCATGCACTGCTGAGGCGGAGAGGAGGAGATGTCAGCCCTTGTGCGAGGG 1819
Qy 391 MetGlnCysMetGluThrHisLysLeuAspProAlaLysGluLeuProGly---TrpGln 409
Db 1820 CTGCAG-----AGGAGCTGAGAGAGACTTCAGAGAGACAGAGGAGCATTTGCAG 1867
Qy 410 -----IleLysGluGlnIleValSerLeu 417
Db 1868 AGTATGTTCCAGAGAACAGAGGAGATCTTAGAGCCACCAAGCAGGAACTCTCTGAGCTG 1927
Qy 418 GluLysAspThrLeuGlnLeuAspLysGluMetGluGluLysAlaArgSerLeuSerLeu 437
Db 1928 CGAATGGAGAGGAGGAGGAGCTTGGAGAGAGATAGAGGTCTTGACAGAGG 1987
Qy 438 MetGluGluAlaAlaLeuAlaLysArgMetTyrAsnGlnGlnIle-----LysArg 454
Db 1988 GAATTAGAGCAGGCCCGAGCTAGTCTCGAGATACTCGCCAGGTTGAGGTGCTCAAGAAG 2047
Qy 455 ProArgLeuSerProMetGlu-----MetProProValThrSerSerSer 469

SEQ ID NO 1242
LENGTH: 2526
TYPE: DNA
ORGANISM: Arabidopsis thaliana
IS-09-938-842A-1242

Alignment Scores:

red. No.: 0.000481 Length: 2526
core: 134.00 Matches: 119
percent Similarity: 34.21% Conservative: 76
est Local Similarity: 20.88% Mismatches: 191
query Match: 4.32% Indels: 184
B: 10 Gaps: 22

IS-09-890-475-1 (1-609) x US-09-938-842A-1242 (1-2526)

```

Y 20 LeuGlnArgHisGlnSerGluGlnArgGluLeuProLysLeuValGluThrGlu 39
b 394 CTTCAAAG---CAAAGTCATCAGAACTTGAAGAACTAAGATTTCAGATGAAGA 450
Y 40 SerThrSerMetAspIleThrIleGln-----SerLysGlnProGln 54
b 451 TACTCCAAAACAGATGCTCTTTATCTCAAGCTTTGTCCAAAACCTCTGTTCTTGAGCAG 510
Y 55 PheLeuLysSerIleAspGluLeu-----AlaAlaPheSer 66
b 511 AACTGAATCTTTGAGGAATATTCGAGAAAGATATCTGAATTAATAATCTGTTTGATA 570
Y 67 ValAlaValGluThrPheLysArgGlnPheAspAspLeuGlnLysHisIleGluSerIle 86
b 571 GTGGCTGAAGAGGAAGAAAGAGTCGAGCAATTCAGATCGAGGAATACCAAGAAAGGTA 630
Y 87 GluAsnAlaIleAspSerLysLeuGluSerAsnGlyValValLeuAlaAlaArgAsnAsn 106
b 631 -----AGTAAATGGAGTCTATCTTTGAACCAATCATCGCGCGAAGAACTCA 675
Y 107 AsnPheHisGlnProMet-----112
b 676 GAGCTTGAAGAACTTTGAGAAATGCTTTGCAGAAAGTGCAGAGCATGAAATTTGGA 735
Y 113 ---LeuSerProProArgAsn-----118
b 736 AATGTGAGTACTAAGCGCGAGCGTTGAACCTACAAGGTCTATTCCAAACATCTCAGTTAAA 795
Y 119 -----AsnValSer 121
b 796 CTAGAAAGACGAGAAATTAAGAACTTGAAGACTTGAAGCAATACAAGTGAATAACTCGAGC 855
Y 122 ValGluThrThrValThrValSerGlnProSerGlnGluIleValProGluThrSerAsn 141
b 856 CTTGAAGCTACTTTAAGCGTTGCAATGAGAAAGAGAGGAGCTTGTCAAGAAATCTTAAT 915
Y 142 LysProGluGlyArgMetCysGluLeuMetCysSerLysGlyLeuArgLysTyrlle 161
b 916 -----GCTGTGATGGAGAGCTGAAATCTCTGAAGAGAGAGACTTGAGAG 960
Y 162 TyrAlaAsnIleSerAspGlnAla-----169
b 961 CAAGCAAGAGAAATACGCAAGCTTACCACCAAGACATAGAGCTTTGAAGCTTTGCATAAA 1020
Y 170 -----LysLeuMetGluGluIleProSerAlaLeuLysLeuAla 182
b 1021 CACTCAGAGCTTAAAGTTCAAAGACAATGAGAGGATTTTAGTAGCAGAGATACAGAGGCC 1080
Y 183 Lys-----GluProAlaLysPheValLeuAspCysIleGlyLysPheTyrlleGln 199
b 1081 AAGTCITTAGCTCAGAAATCAAAGATCTCGAGGAGAAATAAGAGTATAT-----GAA 1134
Y 200 GlyArgAlaPheThrLysGluSerProMetSerSerAlaArgGlnValSerLeuLeu 219
b 1135 GGGAAATTTGCTGAAGCATGTGTCAATCACTCTTTTTCGAAGAAGAACTTGATCAGTCG 1194
Y 220 IleLeuGluSerPheLeuLeuMetProAspArgGlyLysGlyLysValIle-----237

```

```

Db 1195 TCTGCGAGAAATGAGTTACTAGCAGATACAAACCAACAGCTCAAGATCAAGATTCAAGAA 1254
Qy 238 ---GluSerTrpIleLysAspGluAlaGluThrAlaAlaValAlaIleTrpArgLysArgLeu 256
Db 1255 CTTGAAGGATATCTGATCTGAGTCTGAGAGGAAC-----1287
Qy 257 MetThrGluGlyGlyLeuAlaAlaAlaGluLysMetAspAlaArgGlyLeuLeuLeu 276
Db 1288 -----GCAATCGAAAAGTTA-----1302
Qy 277 ValAlaCysPheGlyValProSerAsnPheArgSerThrAspLeuLeuAspLeuIleArg 296
Db 1303 -----AACCAAGAACACAGAAAGCTAAAGACTTGAATACA 1338
Qy 297 MetSerGlySerAsnGluIleAlaGlyAlaLeuLysArgSerGlnPheLeuValProMet 316
Db 1339 AAGTTCAATCCCATGAA-----AACGTAATCGAGGAGCACAAAAGCGAGTTCTTGAA 1392
Qy 317 ValSerGlyIleValGluSerSerIleLysArgGlyMetHisIleGluAlaLeuGluMet 336
Db 1393 GCATCTGGAGTTGCTGATACT-----AGAAAGTGGAGGTAGAA-----1431
Qy 337 ValTyrThrPheGlyMetGluAspLysPheSerAlaAlaLeuValLeuThrSerPheLeu 356
Db 1432 -----GAGCGTTTGTGAAACTTAACACTCTT 1458
Qy 357 LysMetSerLysGluSerPheGluArg-----AlaLysArgLys 369
Db 1459 GAGTCTCAATCAAGAGCTTGAGAAAGAGAAATGGAGATTGGCTGAAGTGAATATAAG 1518
Qy 370 AlaGlnSerProLeuAlaPheLysGluAlaAlaIleThrLysGlnLeuAlaValLeuSer 389
Db 1519 CTGAACCAAGAACTAGCAAACTCAGGGTCAGAGCCGAGATTTTCAGGCTAAGCTCTCT 1578
Qy 390 ValMetGlnCysMetGluThrHisLysLeuAspProAlaLysGluLeuProGlyTrpGln 409
Db 1579 GTT-----TTAGAGCTGAGAAATATCAACAGCTAAAGAACTTCAG-----1620
Qy 410 IleLysGluGlnIleValSerLeuGluLysAspThrLeuGlnLeuAspLysGluMetGlu 429
Db 1621 -----ATAACCATAGAGATCTACAAACAGCTTACTCTGAAAGAGAA 1665
Qy 430 GluLysAlaArgSerLeuSerLeuMetGluGluAlaAlaLeuAlaLysArgMetTyrAsn 449
Db 1666 AGACTACGATCAGATCTCTCCCTTGAGAA-----GAGAAAAACCAAGTCAAT 1716
Qy 450 GlnGlnIleLysArgProArgLeuSerProMetGluMetProProValThrSerSerSer 469
Db 1717 -----GAGATATATCAGAGCCACCAAGATGAG 1743
Qy 470 TyrSerProIleTyrArgAspArgSerPheProSerGlnArgAspAspGln-----487
Db 1744 CTCGTAAGCTTCAAGCAACTCAAGTAGACAAGTCTAAATCCGATGATATGATATCC 1803
Qy 488 -----AspGluIleSerAlaLeuValSer 495
Db 1804 CAAATCGAGAGCTCAGTGCCTTGTGTTGCT 1833

```

RESULT 8

US-10-097-340-42

; Sequence 42, Application US/10097340

; Publication No. US20030087250A1

; GENERAL INFORMATION:

; APPLICANT: John MONAHAN

; APPLICANT: Manjula GANNAVAPURU

; APPLICANT: Sebastian HOERSCH

; APPLICANT: Shubhangi KAMATKAR

; APPLICANT: Steve G. KOVAIS

; APPLICANT: Rachel B. MEYERS

; APPLICANT: Michael MORRISSEY

; APPLICANT: Peter OLANDT

; APPLICANT: Ami SEN

RESULT 6

208	SerProMetSerSerAla	argGlnValSerLeuLeuLeuLeuLeuLeuLeuLeuMet	227
1270	GAGCAGCAAGCAAAATCTAGAGCTGAAGAATATAAACTCTTAGAAGAAAAGCTG-		1323
228	ProAspArgGlyLysGlyLysValLysIleGluSerTrpIleLysAspGluAlaGluThr	247	
1324	-----AAAGGGAAG-----		1350
248	AlaAlaValAlaTrpArgLysArgLeuMetThrGluGlyLeuAlaAlaAlaGluLys	267	
1351	AGTAGTCTGCT-----CATCCAGGCCACCCTGCTTTTTCAGAGAAAAG	1395	
268	MetAspAlaArgGlyLeuLeuLeuValAlaCysPheGlyValProSerAsnPheArg	287	
1396	TATGACAGT-----ATGGTGCAAAGCCTTGAAGATGTTACTGCTCAATTTGAA	1443	
288	Ser-----	ThrAspLeu	291
1444	AGCTATAAAGCGTTAACAGCCAGTCAGATAGAGAATCTTAAGCTGAGAACTCATCATTA	1503	
292	LeuAspLeuIleArgMetSerGlySerAsn-----GluIleAlaGly	305	
1504	CAGGAAAAGGCGCAAGGCTGGGCAAAATGCAGAGGATGTTTCAGCATCAGATTTGCGCA	1563	
306	AlaLeuLysArgSerGlnPheLeuValProMetValSerGlyIleValGluSerSerIle	325	
1564	ACTGAGAGCTCAAAATCAAGAAATATGTAGAGTGCTCTAGATCTGCAGCCAGTCAGCA	1623	
326	LysArgGlyMetHisIleGluAlaLeuGluMetValTyrThrPheGlyMetGluAspLys	345	
1624	CTAAGGAAACAGAGAAATTAAAGMAATCACAGTTCTCTTTCTCAAAAATAACTGAT---	1680	
346	PheSerAlaAlaLeuValLeuThrSerPheLeuLysMetSerLysGluSerPheGluArg	365	
1681	-----TTCAGAAACCAACTCAAGCAACAGAGGAGAGACTTTTAGAAAA	1722	
366	AlaLysArgLysAlaGlnSerProLeuAlaPheLysGluAlaAlaThrLysGlnLeuAla	385	
1723	CAGCTGGAGATGAGAGGAAGGAAAAGCTGMAAAGAAAATAACACAGCAGATTA---	1779	
386	ValLeuSerSerValMetGlnCysMetGluThrHisLysLeuAspProAlaLysGluLeu	405	
1780	-----ACTGAGAGAAAT	1791	
406	ProGlyTrpGln---IleLysGluGlnIleValSerLeuGluLysAsp---ThrLeuGln	423	
1792	ACAAGTGGCGTCTCTCTATGAGAGACTATATAATAAACAAACCTTTTCAGCTACAA	1851	
424	LeuAspLys---GluMetGluGluLysAlaArgSerLeuSerLeuMetGluGluAlaAla	442	
1852	CTAGATGCTTTTGAAGTAGAAAAACAGGCA-----TTGTTGAATGAACATGGT	1899	
443	LeuAlaLysArgMetTyrAsn-----	Gln	450
1900	GCAGCTCAGAAACAGCTAAATAAATAGAGATTCATATGCTAAATATTGGTGCATCAG	1959	
451	GlnIleLysArgProArgLeuSerProMetGluMetProValThrSerSerSerTyr	470	
1960	AATTTGAAACAAAAAATCAAGCATGTTGTGAAGTTGAAAGATGAAATAGCCAACTCAA	2015	
471	SerProIleTyrArgAspArgSerPheProSerGlnArgAspAspGlnAspGluIle	490	
2020	TGGAAAGTATCAAACTCCGCTGTGCAGCTTCTTAATAAAAAACAAAGTGNACAAAAC	2077	
491	SerAlaLeuValSerSerTyrLeuGly-----	ProSerThrSerPhePro	505
2080	CAGAGGAATTGATAAAGTTCTAGTATCAAAACACTTTTGATCCTTCAAAAGGCTTTTCAT	2133	
506	HisArgSerArgArgSerProGluTyrMetValProLeuProHisGly	521	
2140	CATGAAAGTAAAGAAAATTTTCCCTCGAAGACCCCATTAAGAAAGGC	2187	

1210 GATCAGCTTCGATAAAATTACAGCAAAAGGAGGAAACAAGCTGAAGAGCGTGGGTCAAGCAAATG 1268
 208 SerProMetSerAlaArgGlnValSerLeuLeuLeuLeuGluSerPheLeuLeuMet 227
 : : : : :
 1270 GAACAGAACCAAAATCTAGAGCTGAAGAATTAATAACTCTTAGAAGAAAAGCTG----- 1323
 228 ProAspArgGlyIysGlyIysValIysIleGluSerTrpIleLysAspGluAlaGluThr 247
 : : : : :
 1324 -----AAAGGGAG----- 1350
 248 AlaAlaValAlaTrpArgIysArgLeuMetThrGluGlyGlyLeuAlaAlaGluLys 267
 : : : : :
 1351 AGTAGTGCTGCT-----CATACCAGGCCACCCTGCTTTTGCAGAAAAAG 1395
 268 MetAspAlaArgGlyLeuLeuLeuLeuValAlaCysPheGlyValProSerAsnPheArg 287
 : : : : :
 1396 TATGACAGT-----ATGGTGC AAAAGCGCTTGAAGATGTTACTGCTCAATTTGAA 1443
 288 Ser----- 291
 : : : : : ThrAspLeu
 1444 AGCTATAAAGCGTTAACAGCCAGTGAGATAGAAGATCTTAAGCTGAGAACTCATCATTA 1503
 292 LeuAspLeuIleArgMetSerGlySerAsn-----GluIleAlaGly 305
 : : : : :
 1504 CAGGAAAAAGCGGCCAAGGCTGGGAAAAATGCAAGAGATGTCAGCATCATGATTTTGCCA 1563
 306 AlaLeuIysArgSerGlnPheLeuValProMetValSerGlyIleValGluSerIle 325
 : : : : :
 1564 ACTGNAGCTCAATCAAGATATGTAAGGATCTTCTAGATCTGCAGACCAAGTCAGCA 1623
 326 LysArgGlyMetHisIleGluAlaLeuGluMetValTyThrPheGlyMetGluAspLys 345
 : : : : :
 1624 CTAAAGGAAACAGAAAAATTAAAGAAATCACAGTTTCTTTTCTCAAAAATAACTGAT--- 1680
 346 PheSerAlaAlaLeuValLeuThrSerPheLeuLysMetSerLysGluSerPheGluArg 365
 : : : : :
 1681 -----TTCGACAACCACTCAAGCAACAGGAGGAGACATTTAGAAAA 1722
 366 AlaLysArgLysAlaGlnSerProLeuAlaPheIysGluAlaAlaThrIySGlnLeuAla 385
 : : : : :
 1723 CAGCTCGGAAGATGAAGAGGAGGAAAGAGCTGAAAAAGAAAAATCAACACAGCAANTTA--- 1779
 386 ValLeuSerSerValMetGlnCysMetGluThrHisLysLeuAspProAlaLysGluLeu 405
 : : : : :
 1780 -----ACTCAAGAAAATT 1791
 406 ProGlyTrpGln----IleLysGluGlnIleValSerLeuGluLysAsp---ThrLeuGln 423
 : : : : :
 1792 AACCAAGTGGCGTCTCCTCTATGAAGAACTATATAATAAAACAAAACCTTTTCAGCTACAA 1851
 424 LeuAspLys--GluMetGluGlyLysAlaArgSerLeuSerLeuMetGluGluAlaAla 442
 : : : : :
 1852 CTAGATGCTTTTGAAGTAGAAAAACAGCCA-----TTGTTGAATGAACATGGT 1895
 443 LeuAlaLysArgMetTyAsn-----Gln 450
 : : : : :
 1900 GCAGCTCAGGAAACAGCTAAATAAATAAGAGATTATATGCTAAATATTGGGTGCTATCAG 1955
 451 GlnIleLysArgProArgLeuSerProMetGluMetProProValThrSerSerTy 470
 : : : : :
 1960 AATTGAAACAAAAAATCAAGCATGTTGTGAAGTTGAAGATGAAAAATAGCCAACTCAA 2015
 471 SerProIleTyArgAspArgSerPheProSerGlnArgAspAspGlnAspGluIle 490
 : : : : :
 2020 TCGGAAGTAGTACAAATCCGCTGTGCAGCTTCTTAAAAAAAACAAAGTGTAGACAAANCTT 2075
 491 SerAlaLeuValSerSerTyLeuGly-----ProSerThrSerPhePro 505
 : : : : :
 2080 CAACAGGAATTGATAAAGTTCTAGGTATCAACACATTTGTATCCTTCAAAGGCTTTTCAT 2133
 506 HisArgSerArgArgSerProGluTyMetValProLeuProHisGly 521
 : : : : :
 2140 CATCAAGAGTAAAGAAAAATTTTCCCTCGAAGACCCCATTAAAGAAAGGC 2187

```
QY 116 ProArgAsnValSerValGluThrThrValThrValSerGlnProSerGlnGluIle 135
DB 1495 CTCGGATTGGCTTGGAGAGAGAAACCATG---TTGAATAAAAGACAAACAAATT 1551
2Y 136 ValProGluThrSerAsnLysProGluGlyGlyArgMetCysGluLeuMetCysSerLys 155
DB 1552 CAGGATATGGCTGAGAGAG---GGACACAAGCTGGAGATACATACAGCTCAAG 1605
2Y 156 GlyLeuArgLysTyrIleTyrAlaAsnIleSerAspGlnAlaLysLeuMetGluGluIle 175
DB 1606 GACATCTGGATGGAAGGAGCGAAGTTAATGTTCTTCAGAGAAGATTGAAATCTT 1665
QY 176 ProSerAlaLeuLysLeuAlaLysGluProAlaLysPheValLeuAspCysLleGlyLys 195
DB 1666 CAAGAGCAGCTAGA----- 1680
QY 196 PheTyrLeuGlnGlyArgAlaPheThrLysGluSerProMetSerSerAlaArgGln 215
DB 1681 -----GACAAGGAAAGCAGATGAGCAGCTTGAAGAA 1713
QY 216 -----ValSerLeuLeuLeuGlu 222
DB 1714 CGGTCAAATCCTTGACGGCTGACACCACCACTGACACTGCCTTGACAACCTTGGAG 1773
QY 223 SerPheLeuLeuMetProAspArgGlyLysValLysIleGluSerTrpIleLys 242
DB 1774 GAGCCCTTGCAGAGAAAGAGCGGACAATTGAACGCTTAAAG---GAGCAGAGGACAGA 1830
QY 243 AspGluAlaGluThrAlaAlaValAlaTTPArgLysArgLeuMetThrGluGlyGlyLeu 262
DB 1831 GATGAGCGAGAG----- 1842
QY 263 AlaAlaAlaGluLysMetAspAlaArgGlyLeuLeuLeuValAlaCysPheGlyVal 282
DB 1843 ---AAGCAAGAGGAAATTGAT----- 1860
QY 283 ProSerAsnPheArgSerThrAspLeuLeuAsp-----LeuIleArg 296
DB 1861 -----AACTACAAAAA---GATCTTAAAGACTTGAAGGAAAGATCAGCGCTGTGCA 1911
QY 297 MetSerGlySerAnGluIleAlaGlyAlaLeuLysArgSerGlnPheLeuValProMet 316
DB 1912 GCGACCTTTCAGAGAAGAGGCTTCACTTTTGAT-----CTGAAGAGAGAT 1959
QY 317 ValSerGlyIleValGluSerSerIleLysArgGlyMetHisIleGluAlaLeuGluMet 336
DB 1960 GCTTCTTCTGTCATCCTCAGACTGAAAAGAGACTCAGCGCTTAAGACACTAGAGATT 2019
QY 337 ValTyrThrPheGlyMetGluAspLysPheSerAlaAlaLeuValLeuThrSerPheLeu 356
DB 2020 -----GCTTTGGAGAGAGAGAGAGGAGTGTCTGAAAATGGAATCACAATTG 2067
QY 357 Lys-----MetSerLysGluSerPheGluArgAla 366
DB 2068 AAAAGGCACATGAGCGCAGCATTTGAAGCCAGAGCCAGTCCAGAGATGATGACCGAATA 2127
QY 367 LysArgLysAlaGlnSerProLeuAlaPheLysGluAlaAlaThrLysGlnLeuAlaVal 386
DB 2128 CAGCACTTGGAGAGAGAGATCACAGGTACAAAGATGAATCTAGCAAGCCCGAGCAGAA 2187
QY 387 LeuSerSerValMetGlnCysMet-----GluThrHisLysLeuAspProAlaLys 403
DB 2188 GTTGATCGACTCTTAGAAATCTTGAAGAGAGGTGGAATGAGAAGATGACAAGATTAAG 2247
QY 404 Glu-----LeuProGlyTrpGlnIleLysGluGln-----Ile 414
DB 2248 AAGATAGCTGAGTTGGAAGATCTCACTCAGGCAAGTGAAGACCGAATAAGAGGTA 2307
QY 415 ValSerLeuGluLysAspThrLeuGlnLeuAspLysGluMetGluGluLysAlaArgSer 434
DB 2308 GCAAATCTGAAG-----CACAGGAAACAGGTGGAAGAAAGAGAGAT 2349
```

```
QY 435 LeuSerLeuMetGluGluAla 441
DB 2350 GCACAAATGTTAGAGGAGCG 2370
RESULT 5
US-09-960-253-159
; Sequence 159, Application US/09960253
; Patent No. US20020123619A1
; GENERAL INFORMATION:
; APPLICANT: Benson, Darin R.
; APPLICANT: Mohamath, Raodoh
; APPLICANT: Lodes, Michael J.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.556
; CURRENT APPLICATION NUMBER: US/09/960,253
; CURRENT FILING DATE: 2001-09-20
; NUMBER OF SEQ ID NOS: 187
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 159
; LENGTH: 2756
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-960-253-159
Alignment Scores:
Pred. No.: 4,52e-05 Length: 2756
Score: 144.00 Matches: 109
Percent Similarity: 35.97% Conservative: 91
Best Local Similarity: 19.60% Mismatches: 226
Query Match: 4.64% Indels: 130
DB: 20 Gaps: 20
US-09-890-475-1 (1-609) x US-09-960-253-159 (1-2756)
QY 34 LysIleValGluThrGluSerThrSerMetAspIleThrIleGlyGlnSerLysGlnPro 53
DB 706 AAATCTGTTTCAATAGAGAGAAAGAAAGATTGAT-----GAAAAATCTCAACAGAA 756
QY 54 GlnPheLeuLysSerIleAspGluLeuAlaAlaPheSerValAlaValGluThrPheLys 73
DB 757 AAATCTTGGATACATCGAAGAAATAGTTGCTTCAGATCAAGTGGAAAAATACAA 816
QY 74 ArgGlnPheAspLeuGlnLysHisIleGluSerIleGluAsnAlaIleAspSer--- 92
DB 817 CTAGATATTCGCCAGTTAGAGAAAAATTTGAAGAGAGAAAGATGATGAAATTTAAGCCTT 876
QY 93 -----LysLeuGluSerAsnGlyValValLeuAla----- 102
DB 877 AAGCAGTCTCTTGAGGACAATATTTATTTATTTATCTAAACAAGTAGAAGATCTAAATGTG 936
QY 103 -----AlaArgAsnAsnAsnPheHisGln 110
DB 937 AAATGTCAGCTGTTGAAACAGAAAAAGAACACCATGTCTCAACAGGAATAGAGAACACAA 996
QY 111 ProMetLeuSerProProArgAsnAsnValSerValGluThrThrValThrValSerGln 130
DB 997 GAAATCTTAATGTCAGAGATGCAAAAC-----TTAGAACAGAGATTTATCTTGAACAA 1050
QY 131 ProSerGlnIleValProGluThrSerAsnLysProGluGlyGlyArgMetCysGlu 150
DB 1051 CGGGAACATGAAAGCTTCAACAAAAAGAAATTTACAAATTTGATTCTCTCGCAACAGAG 1110
QY 151 LeuMetCysSerLysGlyLeuArgLysTyrIleTyrAlaAsnIleSerAspGlnAlaLys 170
DB 1111 AAAGATTAATCTTCGAGTCTTCATCAGAGCTC-----TGTTCT 1149
QY 171 LeuMetGluGluIleProSerAlaLeuLysLeuAlaLysGluProAlaLysPheValLeu 190
DB 1150 TTTCAAGAGGAAATGTTTAAAGAGAGAAATCTGTTTGAAGAGAAATTTAAGCAACACTG 1209
QY 191 AspCysIleGlyLysPheTyrLeu-----GlnGlyArgAlaPheThrLysGlu 207
```

y	297	tSerGlySerAsnGluIleAlaGlyAlaLeuLysArgSerGlnPheLeuValProMetVa	317
b	2192	ACAGCATCAAACTGAGCTCACTAAACAGCTGAATATATAAAGCGAAGA	2240
y	317	lSerGlyIleValGluSerSerIleIysArgGlyMethIstleGluAlaLeuGluMetVa	337
b	2241	---GAACGAGAACTTAAGACGA---AAGCATGTC---	2267
y	337	lTyrThrPheGlyMetGluAspIysPheSerAlaAlaLeuValLeuThrSerPheLeuIy	357
b	2268	---ATGGAAGTTTCGACACACAGCGCTAAGAGTTTGAAGTCTAAAGAACTCCA	2314
y	357	sMetSerLysGluSerPheGluArgAlaIysArgIysAlaGlnSerProLeuIa	375
b	2315	AATAAAAGACAGTTTCAGGATACCTGCGAAATCCAAACCCAGACAGTACAAAGCATTAAG	2374
y	376	---PheIy	377
b	2375	AAATCACCTGCTGGAGACTACCAAAAGAGTGAGCACAAAGCTGTCTCTGAAACGGCTCAA	2434
y	377	sGluAlaAlaThrLysGlnLeuAlaValLeuSer	392
b	2435	GGAGGAACACAGCCCGAAATTAGCTATCTTGCTGAGCAGTATGATCACAGCATTAATGA	2494
y	392	nCysMetGluThrHis	410
b	2495	AATGCTCTCCACACAGCCCTGGGTTTGATGATGACGACAGACAGAGTCCCGAGTTT	2554
y	410	eLysGluGlnIle	417
b	2555	GAAGATCGAGCTGCAGCAGCAACTGGAGCTGTGTAAATCGCTATCAGAGCAAAATCAAGT	2614
y	417	uGluLysAspThrLeuGlnLeuAspLysGluMetGluGluLysAlaArgSerLeuSerLe	437
b	2615	GCAAAGCTTGAGCA---CAACATGATCGAGAGCTTCGGAGCTTGACAGAGCGTCTCCCT	2671
y	437	u	449
b	2672	CCGGAGGGCACTCTTAGAACAAAGATTGAAGAGAGATGTGGCT---TTGCAGAA	2725
y	449	nGlnGlnIleIysArgProArgLeuSerProMetGluMetProProValThrSerSerSe	469
b	2726	TGAGCGCAGACAGCAATACGA	2747
y	469	rTySerProIleTyArgAspArgSerPheProSerGlnArgAspAspGlnAspGl	489
b	2748	---AGCCTGTTTGGAAAGCTCAAGCCAGACAGA	2773
y	489	uIleSerAlaLeuValSerSerTyLeuGlyProSerThrSerPheProHisArgSerAr	509
b	2774	GATTGAAGCTTTTGACTCTGAACAGCATGAGACTAGGTTTTAGTAAATATGGTCTCTTCTAA	2833
y	509	gArgSerProGluTyMetValProLeuProHisIsglyGlyLeuGlyArgSerValTyAl	529
b	2834	TCCTCCCTCGAG	2848
y	529	aTyGluHisLeuAlaProIle	543
b	2849	ATTGAGCCACAGCTACCCGGAGGCTCTGTTGGTTCACACAAACCTACTGGGGGTCCAGG	2908
y	544	---HisArgLeuHisArgGlnTySerProSerLeuValHisIsglyGlnArgHisProLe	562
b	2909	ACCTCCTGGGGTCACTCCCATGGTGGGCCCAACACAGCTTGGGGC	2962
y	562	uGlnTySerPro	577
b	2963	GCAAAGTGGACCCACAGCCATGGGGTCACTCCCTCAGGGCCAAATGCAAGGGGTACCTCGAGG	3022
y	577	ylleGlnArgValTyArgHisSerProSerGluGluArgTyLeu	594
b	3023	TAGCAGTATGGAGTCCGCAATAGCCCCCAGGCTCTTGAGCGCGGACAGCTTCTGCGGGGACG	3082
y	594	uSerAsnGlnArgSerProArgSerAsnSer	604

```

Db      3083 GACAGAGCAGGGCATGAGCAGAGACGAGGT 3113
      ... ||| ||||| ||||| ||||| |||||
RESULT 4
US-09-893-519A-146
; Sequence 146, Application US/09893519A
; Publication No. US20030027243A1
; GENERAL INFORMATION:
; APPLICANT: ANADYS PHARMACEUTICALS, INC.
; APPLICANT: THOMPSON, Craig
; APPLICANT: MOORE, Jeffrey
; APPLICANT: BUURMAN, Ed T.
; APPLICANT: BRADLEY, John
; APPLICANT: DESILVA, Thamara
; APPLICANT: HARRIS, Sandra
; APPLICANT: KONARNISKY, Svetlana
; APPLICANT: MENDILLO, Marc
; APPLICANT: MOORE, Daniel
; APPLICANT: MCCOY, Melissa
; APPLICANT: SANDERSON, Karen
; APPLICANT: HAO, Tariq
; APPLICANT: ZHU, Shuhao
; APPLICANT: LONG, Fan
; APPLICANT: DAVIDOV, Eugene
; TITLE OF INVENTION: ANTIFUNGAL COMPOUNDS AND METHODS OF USE
; FILE REFERENCE: 0342/1G548-US2
; CURRENT APPLICATION NUMBER: US/09/893,519A
; CURRENT FILING DATE: 2001-06-28
; PRIOR APPLICATION NUMBER: US 60/215,164
; PRIOR FILING DATE: 2000-06-29
; PRIOR APPLICATION NUMBER: US 60/224,457
; PRIOR FILING DATE: 2000-08-10
; NUMBER OF SEQ ID NOS: 146
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 146
; LENGTH: 2760
; TYPE: DNA
; ORGANISM: Homo sapiens
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: Human Genbank/AB015617.1
; DATABASE ENTRY DATE: 2000-01-06
; RELEVANT RESIDUES: (1)..(2760)
US-09-893-519A-146

Alignment Scores:
Pred. No.: 3,11e-05 Length: 2760
Score: 145.50 Matches: 95
Percent Similarity: 37.69% Conservative: 81
Best Local Similarity: 20.34% Mismatches: 164
Query Match: 4.69% Indels: 127
DB: 11 Gaps: 20

US-09-890-475-1 (1-609) x US-09-893-519A-146 (1-2760)
QY      16 AlaAsnProLeuLeuGlnArgHisGlnSerGluGlnArgArgGluLeuProLysile 35
Db      1228 TCGAATGGTGCTTTGACTACTGAGGAAGGGAAGAGAAATGAAGCAATG---GAAGTG 1284
QY      36 ValGluThrGluSerThrSerMetAspIleThrIleGlyGlnSerLysGlnProGlnPhe 55
Db      1285 TATCGGAGCCATTCTAAATTTATGAAAAAATAAGATTGGCCAGGTGAAACAG---GAGCTG 1341
QY      56 LeuLysSerIleAspGluLeuAlaAlaPheSerValAlaValGluThrPheLysArgGln 75
Db      1342 TCCGAAGAGCACAGAACTACTCGCCCTGCAGACAAAGCTAGAAACTCACAAPCCAG 1401
QY      76 PheAspAspLeuGlnLysHisIleGluSerIleGluAsnAlaIleAspSerLysLeuGlu 95
Db      1402 TTCTCAGATAGTAAACAGCACTTGAAGTGTGTGAAGGAGTCTTGTGCTAAG---GAG 1458
QY      96 SerAsnGlyValValLeuAlaAlaArgAsnAsnPheHisGlnProMetLeuSerPro 115
Db      1459 CAGAGGGGTGTCATCTCTG-----CAGACTGAGGTGATGCT 1494

```

QY 529 aTyrGluHisLeuAlaProAsn-----SerTyrSerProGlyHisGly----- 543
DB 2053 ATTACGCCACAGCTACCCGGGAGCTTCTGGTGTGCACACACCTACTGGGGTCCAGG 2112
QY 544 ----HisArgLeuHisArgGlnTyrSerProSerLeuValHisGlyGlnArgHisProLe 562
DB 2113 ACTCACTGGGTGCATCCCATGGGTGGCCACCAAGCTTGGGGC-----CATCCAA 2166
QY 562 uGlnTyrSerPro-----ProIleHisGlyGlnGlnGlnLeuProTyrG 577
DB 2167 GCAAGGTGACCCCGCCATGGGGTCCACCTTCAGGGCCATGCAGGGGTACCTCGAGG 2226
QY 577 yIleGlnArgValTyrArgHisSerProSerGluGluArgTyrLeu-----GlyLe 594
DB 2227 TAGCAGTGGGAGTCCGCAATAGCCCCAGGCTCTGAGGGGACAGCTTCTGGGGGACG 2286
QY 594 uSerAsnGlnArgSerProArgSerAsnSer 604
DB 2287 GACGGAGCAGGGCATGAGCAGAAGCAGGAGT 2317

RESULT 3

US-10-177-293-253
; Sequence 253, Application US/10177293
; Publication No. US20030124128A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Glatt, Karen
; APPLICANT: Zhao, Xumei
; APPLICANT: Gannavarpu, Manjula
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Mertens, Maureen
; APPLICANT: Myer, Vic
; APPLICANT: Wang, Youzheng
; APPLICANT: Xu, Yongyao
; APPLICANT: Hoerschen, Sebastian
; APPLICANT: Monahan, John
; APPLICANT: Meyers, Rachel E.
; APPLICANT: Bast Jr., Robert C.
; APPLICANT: Hortobagyi, Gabriel N.
; APPLICANT: Pusztai, Lajos
; APPLICANT: Meric, Funda
; APPLICANT: Sahin, Aysegul
; APPLICANT: Mills, Gordon B.
; TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR IDENTIFICATION, ASSESSMENT,
; PREVENTION, AND THERAPY OF BREAST CANCER
; FILE REFERENCE: MRI-038
; CURRENT APPLICATION NUMBER: US/10177,293
; CURRENT FILING DATE: 2002-06-21
; PRIOR APPLICATION NUMBER: US 60/299,887
; PRIOR FILING DATE: 2001-06-21
; PRIOR APPLICATION NUMBER: US 60/301,572
; PRIOR FILING DATE: 2001-06-27
; PRIOR APPLICATION NUMBER: US 60/306,501
; PRIOR FILING DATE: 2001-07-18
; PRIOR APPLICATION NUMBER: US 60/325,002
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US 60/362,585
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: US 60/xxx,xxx
; PRIOR FILING DATE: 2002-05-14
; NUMBER OF SEQ ID NOS: 506
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 253
; LENGTH: 4620
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-177-293-253

Alignment Scores:
Pred. No.: 3.26e-05 Length: 4620
Score: 148.50 Matches: 150
Percent Similarity: 34.88% Conservative: 98
Best Local Similarity: 21.10% Mismatches: 227

Query Match: 4.78% Indels: 237
DB: 14 Gaps: 35
US-09-890-475-1 (1-609) x US-10-177-293-253 (1-4620)
QY 3 AsnTyr-----ProProThrValAlaAlaGlnProThrThrThrAlaAsn 17
DB 1365 AATTACAGAGAGAGGAGATCTTAGAACAGAGCATCAGATCCAACTCCA----- 1418
QY 18 ProLeuLeuGlnArgHisGlnSerGluGlnArgArgGluLeuProLysIleValGlu 37
DB 1419 CCCCAAGTATCTCGTCACAAATCACACTATCGTAATCGAGAACACATTTGCTACTACGG 1478
QY 38 ThrGluSerThrSerMetAspIleThrIleGlyGlnSerLysGlnProGlnPheLeuLys 57
DB 1479 ACAGCATCAGT-G-----GTTACGAGGCAATGCAAGAACATGAG 1517
QY 58 SerIleAspGluLeuAlaAlaPheSerValAlaValGluThrPheLysArgGlnPheAsp 77
DB 1518 CAGGACTCTGAGCTTAGAGAACAAATGCTGGCTATTAAGCGAATGAGGCGACAACAT-- 1574
QY 78 AspLeuGlnLysHisIleGluSerIleGluAsnAlaIleAspSerLysLeuGluSerAsn 97
DB 1575 -----CAAAAGCACTGATGACTCTGAAAACCAAGCTAAAGGTGAGATGGAATGAACAT 1628
QY 98 GlyValValLeuAlaAlaArgAsnAsnAsnPheHisGlnProMetLeuSerProProArg 117
DB 1629 CGCTTCAGATTA-----GACAAAGATCTTGAACCTCAGCGT 1664
QY 118 AsnAsnValSerValGluThrThrValThrValSerGlnProSerGlnGluIleValPro 137
DB 1665 AACAAATTTTCTGCAGAAATGGAGAACTTATCAAGAAACACCACCGCTGTATATGGAGAAA 1724
QY 138 GluThr-----SerAsnLysProGluGlyArgMetCysGluLeuMetCysSer 154
DB 1725 GAGGCTAAGTATGTCATGAAAGAA-----GAG 1751
QY 155 LysGlyLeuArgLysTyrIleTyrAlaAsnIleSerAspGlnAlaLysLeuMetGluGlu 174
DB 1752 AAAAAATTTTCAGCAACATATT-----CAGGCCCAACAGAGAAAGAA 1793
QY 175 IleProSerAlaLeu-----LysLeuAlaLysGluProAlaLys 187
DB 1794 CTGAATAGTTTCTCGAGTCCAGAAAGAGAGATATAACTTCCAAAGAGCAGCTTAAA 1853
QY 188 PheValLeuAspCysIleGlyLysPheTyrLeuGlnArgArg-----AlaPhe 204
DB 1854 GAGGAGCTAAAT--GAAACACCAGAG-TACCCCAAAAAAGAAAAACAGGAGTGGCTTTC 1909
QY 205 ThrLysGluSer-----ProMetSerSer 212
DB 1910 AAAGCAGAGAGGAGATATACAGCATTTCCAGCAGAGAGAGAGTAACTTCTTCGACG 1969
QY 213 AlaArgGlnValSerLeuLeuIleLeuGluSerPheLeuLeuMetProAspArgLysLys 232
DB 1970 TCAAGACAAATA-----CCTAGAGCTGGAATG 1996
QY 233 GlyLysValLysIleGluSerTrpIleLysaspGluAlaGluThrAlaAlaValAlaTrp 252
DB 1997 CCCTCGCTTCAAGAGAGAAAT-----GTTACTTGG 2026
QY 253 ArgLysArgLeuMetThr-----GluGlyGlyLeuAlaAlaAlaGluLysMetAsp 269
DB 2027 CGCTATAACTTAGACGAGGACCTTGTTCAGGGAGGAGTT-----AAACAAAAGACAGAC 2080
QY 270 AlaArgGlyLeuLeuLeuValAlaCysPheGlyValProSer----- 284
DB 2081 TCAGAAAGGACTTAGA-----GCATGCCATGTACTCCGACAGCATGAATCTATGCA 2131
QY 285 -----AsnPheArgSerThrAspLeuLeu-AspLeuIleArgMe 297
DB 2132 AGAAGTGGAGTTCGGCCACCTCAACACAAATTCAGAGAGATCGCTGTGAGTTGATCAGATT 2191

CURRENT APPLICATION NUMBER: US/09/291,417A
 CURRENT FILING DATE: 1999-04-13
 EARLIER APPLICATION NUMBER: US 60/081,784
 EARLIER FILING DATE: 1998-04-14
 NUMBER OF SEQ ID NOS: 147
 SOFTWARE: FastSeq for Windows Version 3.0
 SEQ ID NO 20
 LENGTH: 3824
 TYPE: DNA
 ORGANISM: Mammalian (Human) SULU3
 S-09-291-417-20

Alignment Scores:
 Seq. No.: 2,43e-05 Length: 3824
 Score: 148.50 Matches: 150
 Percent Similarity: 34.88% Conservative: 98
 Best Local Similarity: 21.10% Mismatches: 227
 Query Match: 4.78% Indels: 237
 Gaps: 35

US-09-890-475-1 (1-609) x US-09-291-417-20 (1-3824)

```

Y 3 AsnTyr-----PropThrValAlaAlaGlnProThrThrThrAlaAsn 17
b 569 AATTACAGAGAGAGGAGAGTCTAGAACAGAGCATCAGATCCCAATCTCCA----- 622
Y 18 ProLeuGlnArgHisGlnSerGluGlnArgArgGluLeuProLysIleValGlu 37
b 623 CCCAAGTATCTCGTCAAAATGACATATGCTGAAATCGAGAGACCTTCTCTACTACGG 682
Y 38 ThrGluSerThrSerMetAspIleThrIleGlyGlnSerLysGlnProGlnPheLeuLys 57
b 683 ACAGCATCACTG-----GTTACGAGGCAATGCAAGAACATGAG 721
Y 58 SerIleAspGluLeuAlaAlaPheSerValAlaValGluThrPheLysArgGlnPheAsp 77
b 722 CAGGACTCTGAGCTTAGAGAACAAATGCTCGGTATAGCGAATAGAGGCGACAACAT--- 778
Y 78 AspLeuGlnLysHisIleGluSerIleGluAsnAlaIleAspSerLysLeuGluSerAsn 97
b 779 -----CAAAAGCAACTGTAGTACTCTGGAAAAACAAGCTAAAGGCTGAGATGAACAT 832
Y 98 GlyValValLeuAlaAlaArgAsnAsnPheHisGlnProMetLeuSerProProArg 117
b 833 CGCCTCAGATTA-----GACAAAGATCTTGAAGACTCAGCGT 868
Y 118 AsnAsnValSerValGluThrThrValThrValSerGlnProSerGlnGluIleValPro 137
b 869 AACCAATTTGCTGAGAAATGGAGAAACTTATCAAGAAACACAGGCTGCCATGGAGAAA 928
Y 138 GluThr-----SerAsnLysProGluGlyGlyArgMetCysGluLeuMetCysSer 154
b 929 GAGGCTAAAGTGATGCCAATGAA-----GAG 955
Y 155 LysGlyLeuArgLysTyrIleTyrAlaAsnIleSerAspGlnAlaLysLeuMetGluGlu 174
b 956 AAAAAATTCAGCAACATATT-----CAGGCCCAACAGAGAAGAA 997
Y 175 IleProSerAlaLeu-----LysLeuAlaLysGluProAlaLys 187
b 998 CTGAATAGTTTCTCGAGTCCCAAGAGAGAGTATAACTTCGAAAAGAGCAGCTTAAA 1057
Y 188 PheValLeuAspCysIleGlyPheTyrLeuGlnGlyArg-----AlaPhe 204
b 1058 GAGGAGCTTAAT--GAAACACAGAG-TACCCCCAAAAAGAAAAACAGAGTGGCTTTC 1113
Y 205 ThrLysGluSer-----ProMetSerSer 212
b 1114 AAAGCAAGAGAGATATACAGCATTTCCAGACAGAGAGAGACTTAACCTTCTTCGAGC 1173
Y 213 AlaArgGlnValSerLeuLeuLeuGluSerPheLeuLeuMetProAspArgGlyLys 232
b 1174 TCMAAGACAATA-----CCTAGAGTGGAAATG 1200

```

```

QY 233 GlyLysValLysIleGluSerTrpIleLysAspGluAlaGluThrAlaAlaValAlaTrp 252
Db 1201 CGTCGCTTCAAGAGAAGAA-----GTTACTTGG 1230
QY 253 ArgLysArgLeuMetThr-----GluGlyGlyLeuAlaAlaAlaGluLysMetAsp 269
Db 1231 GCGTCATAACTTAGAGCAGGACCTTGTCCAGGAGAGTT-----AAACAAAGACAGAC 1284
QY 270 AlaArgGlyLeuLeuLeuValAlaCysPheGlyValProSer----- 284
Db 1285 TCAGAGGAGCTTAGA-----GCATGCCATGCTACTCCGACAGCATGAATCTATGCA 1335
QY 285 -----AsnPheArgSerThrAspLeuLeuAspLeuIleArgMe 297
Db 1336 AGAAGCTGGAGTTCGCCACCTCAACACATTCAGAGATGCGCTGTGAGTTGATCAGATT 1395
QY 297 tSerGlySerAsnGluIleAlaGlyAlaLysArgSerGlnPheLeuValProMetVa 317
Db 1396 ACAGCATCAAACTGAGCTCACTAACGAGCTGGAATATATAAGCGAAGA----- 1444
QY 317 lSerGlyIleValGluSerSerIleLysArgGlyMethIleGluAlaLeuGluMetVa 337
Db 1445 -----GAACGAGAACTAAGACGA---AAGCATGTCT----- 1471
QY 337 lTyrThrPheGlyMetGluAspLysPheSerAlaAlaLeuValLeuThrSerPheLeuLys 357
Db 1472 -----ATGGAAGTTCGACACACAGCTAAGAGTTTGAAGTCTAAAGAACTCCA 1518
QY 357 sMetSerLysGluSerPheGluArgAlaLysArgLysAlaGlnSerProLeuAla----- 375
Db 1519 AATAAAAAGCAGATTTCAGGATACCTGCAAAATCCAAACCCAGACAGTCAAAAGCATTAAG 1578
QY 376 -----PheLys 377
Db 1579 AATCACCCTGCTGGAGACTACACCAAGAGTGCAGACAAAGCTGTTCTGAAACGCTCAA 1638
QY 377 sGluAlaAlaThrLysGlnLeuAlaValLeuSer-----SerValMetG1 392
Db 1639 GGAGGAACAGACCGGAAATTAGCTATCTTGCTGAGCAGCATGATCATCAGCATTAATGA 1698
QY 392 nCysMetGluThrHis-----LysLeuAspProAlaLysGluLeuProGlyTrpGlnI1 410
Db 1699 AATGCTCTCCACCAAGCCCTCGGTTTGATTAAGCAGACAGGAGCAGAGTGCAGGTTT 1758
QY 410 eLysGluGlnIle-----ValSerLe 417
Db 1759 GAAGATGACGCTGCAGCAGAACTGGAGCTGTTGAATGCGTATCAGAGCAAAATCAAGAT 1818
QY 417 uGluLysAspThrLeuGlnLeuAspLysGluMetGluGluLysAlaArgSerLeuSerLe 437
Db 1819 GCAAGCTGAGGCA---CAACATGATCGAGAGCTTCGAGAGCTTGAACAGAGGCTCCCT 1875
QY 437 u-----MetGluGluAlaAlaLeuAlaLysArgMetTyAs 449
Db 1876 CCGGAGGCGCACTCTTAGAACAAAGATTCAAGAGAGAGATTGGCT-----TTGCAGAA 1929
QY 449 nGlnGlnIleLysArgProArgLeuSerProMetGluMetProProValThrSerSerSe 469
Db 1930 TGAGCGCACAGAACGAATACGA----- 1951
QY 469 rTyrSerProIleTyrArgAspArgSerPheProSerGlnArgAspAspGlnAspG1 489
Db 1952 -----AGCCTGTTGGAACGCTCAAGCCAGCAGAGA 1977
QY 489 uIleSerAlaLeuValSerSerTyrLeuGlyProSerThrSerPheProHisArgSerAr 509
Db 1978 GATTGAAGCTTTTGAAGCTGAAAGCATGAGACTAGGCTTTTGTAGTATATGCTTCTTAA 2037
QY 509 gArgSerProGluTyrMetValProLeuProHisGlyGlyLeuGlyArgSerValTyAl 529
Db 2038 TCTCTCCCTCGAG-----GC 2052

```


GenCore version 5.1.6
Copyright (C) 1993 - 2003 CompuGen Ltd.

M protein - nucleic search, using frame_plus_p2n model

un on: August 25, 2003, 12:30:10 ; Search time 384 Seconds

(without alignments)

3565.447 Million cell updates/sec

file: US-09-890-475-1

effect score: 3104

sequence: 1 MGNYPPTVAAQPTTANPL.....RYGLSNQSPRSNSLDPK 609

coring table: BLOSUM62

Xgapop 10.0, Xgapext 0.5

Fgapop 10.0, Fgapext 0.5

Fgapop 6.0, Fgapext 7.0

Delop 6.0, Delext 7.0

searched: 1517243 seqs, 1124081882 residues

total number of hits satisfying chosen parameters: 3034486

inimum DB seq length: 0

aximum DB seq length: 2000000000

ost-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

MODEL=frame+ p2n.model -DRV=xlh

Q=/cgn2_1/USPTO.spool/US09890475/runat_14082003_090354_19703/app_query.fasta_1.775

DB=Published Applications NA -OFMT=fastcap -SUFFIX=rnpb -MINMATCH=0.1

LOPCL=0 -LOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blomsum62

TRANS-human40.cdd -LIST=45 -DICALIGN=200 -THR SCORE=pct -THR MAX=100

THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0

MAXLEN=200000000 -USER=US09890475@cgn_1_1238@runat_14082003_090354_19703

NCPU=6 -ICPU=3 -NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100

LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5

FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

database :

Published Applications NA:

1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq*

2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq*

3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq*

4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq*

5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq*

6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq*

7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq*

8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq*

9: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq*

10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq*

11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq*

12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq*

13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq*

14: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq*

15: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq*

16: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq*

17: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

result	Query	Match	Length	DB	ID	Description
1	234.5	7.6	1524	10	US-09-938-842A-917	Sequence 917, App

2	148.5	4.8	3824	11	US-09-291-417-20	Sequence 20, Appl
3	148.5	4.8	4620	14	US-10-177-293-253	Sequence 253, App
4	145.5	4.7	2760	11	US-09-893-519A-146	Sequence 146, App
5	144	4.6	2756	10	US-09-960-253-159	Sequence 159, App
6	140	4.5	3718	10	US-09-917-800A-1471	Sequence 1471, Ap
7	134	4.3	2526	10	US-09-938-842A-1242	Sequence 1242, Ap
8	132.5	4.3	5142	14	US-10-097-340-42	Sequence 42, Appl
9	131	4.2	9775	9	US-09-751-962-1	Sequence 1, Appli
10	130.5	4.2	7393	14	US-10-037-270-372	Sequence 372, App
11	129	4.2	3324	14	US-10-037-270-1020	Sequence 1020, Ap
12	129	4.2	9549	14	US-10-204-887-31	Sequence 31, Appl
13	129	4.2	3309400	10	US-09-738-626-1	Sequence 1, Appli
14	127.5	4.1	10432	10	US-09-919-172-97	Sequence 97, Appl
15	127.5	4.1	10432	11	US-09-919-039-20	Sequence 20, Appl
16	127.5	4.1	10432	14	US-10-084-817-349	Sequence 349, App
17	127.5	4.1	12515	11	US-09-918-624B-48	Sequence 48, Appl
18	127.5	4.1	12632	13	US-10-071-766-49	Sequence 49, Appl
19	126.5	4.1	9934	9	US-09-751-962-2	Sequence 2, Appli
20	124.5	4.0	3284	14	US-10-128-714-7305	Sequence 7305, Ap
21	124.5	4.0	3379	14	US-10-128-714-6305	Sequence 6305, Ap
22	124.5	4.0	5379	14	US-10-128-714-5305	Sequence 5305, Ap
23	124	4.0	1278	14	US-10-156-761-5245	Sequence 5245, Ap
24	124	4.0	9025608	14	US-10-156-761-1	Sequence 1, Appli
25	122.5	3.9	6604	10	US-09-880-107-1748	Sequence 1748, Ap
26	122.5	3.9	7100	10	US-09-932-183A-1	Sequence 1, Appli
27	121	3.9	3811	14	US-10-177-293-105	Sequence 105, App
28	121	3.9	14800	10	US-09-954-456-1601	Sequence 1601, Ap
29	120	3.9	11220	10	US-09-861-289-32	Sequence 32, Appl
30	120	3.9	11220	10	US-09-860-846-32	Sequence 32, Appl
31	120	3.9	11220	11	US-09-988-384B-32	Sequence 32, Appl
32	120	3.9	11220	11	US-09-836-821-32	Sequence 32, Appl
33	120	3.9	36778	10	US-09-861-289-5	Sequence 5, Appli
34	120	3.9	36778	10	US-09-860-846-5	Sequence 5, Appli
35	120	3.9	36778	11	US-09-836-821-5	Sequence 5, Appli
36	120	3.9	37948	11	US-09-988-384B-5	Sequence 3717, Ap
37	119	3.8	3075	10	US-09-880-107-3717	Sequence 102, App
38	119	3.8	3080	14	US-10-171-581-102	Sequence 1921, App
39	119	3.8	7869	10	US-09-954-456-1921	Sequence 297, App
40	119	3.8	7923	10	US-09-864-864-297	Sequence 11, Appl
41	119	3.8	7923	14	US-10-239-058-11	Sequence 1602, Ap
42	118.5	3.8	3388	10	US-09-954-456-1602	Sequence 245, App
43	118.5	3.8	3388	10	US-09-967-768A-245	Sequence 988, App
44	118.5	3.8	3388	10	US-09-954-531-988	Sequence 1382, Ap
45	118.5	3.8	3388	10	US-09-954-531-1382	

ALIGNMENTS

RESULT 1

US-09-938-842A-917

; Sequence 917, Application US/09938842A

; Patent No. US20020160378A1

; GENERAL INFORMATION:

; APPLICANT: Harper, Jeff

; APPLICANT: Kieps, Joel

; APPLICANT: Wang, Xun

; APPLICANT: Zhu, Tong

; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING

; FILE REFERENCE: SAME, AND METHODS OF USE

; FILE REFERENCE: SCRIPT300-3

; CURRENT APPLICATION NUMBER: US/09/938,842A

; CURRENT FILING DATE: 2001-08-24

; PRIOR APPLICATION NUMBER: US 60/227,866

; PRIOR FILING DATE: 2000-08-24

; PRIOR APPLICATION NUMBER: US 60/264,647

; PRIOR FILING DATE: 2001-01-16

; PRIOR APPLICATION NUMBER: US 60/300,111

; PRIOR FILING DATE: 2001-06-22

; NUMBER OF SEQ ID NOS: 5379

; SEQ ID NO 917

; LENGTH: 1524

; TYPE: DNA

; ORGANISM: Arabidopsis thaliana

3790 AAGCTGGAGGAGCTCGCTCTGTCAGGACAGAGACAGCCAGCAACAGTGCAGAGCTG 3849
 140 SerAsnLysPro---GluGlyGlyArgMetCys----- 149
 3850 CAGAACCCAGCTCTGCTCTCGGAGGAGGTGTCAGAGCCCTCCGGAGGAGGCTGAGAA 3909
 150 GluLeuMetCysSerLysGlyLeuArgLysTyrIleTyrAlaAsnIleSerAspGlnAla 169
 3910 CAGCGGTGGCTTCAGAGAACCTGCGCAG-----GAGTGAACCTCACAGGCT 3957
 170 LysLeuMetGluGluIleProSerAlaLeuLysGlu----- 184
 3958 GAGCGTCGGAGGAGCTGGCCAGAAATTGAAGCGTGGCAGGAGAGTCTTCCAGAA 4017
 184 ----- 184
 4018 GAGCAGGCCCTCTCCACCTCGAGCTCGAGCACACAGCAGCCCTCGTGTGAGTGAG 4077
 185 -----ProAlaLysPheValLeuAupCysIleGlyPheTyrLeuGlnGlyArg 202
 4078 CTGCTGCAGCTAAGCACTCTGCCAG-----CAGTGCAGCCGAGCAG 4122
 203 AlaPheThrLysGluSerProMetSerSerAlaArgGlnValSerLeuLeuIle----- 220
 4123 GCCGCTCCGAGAAACCCAGCTGAGAGCTGGAGCAGAGCAAGCAGCCGCTGGGGGA 4182
 221 LeuGluSerPheLeuMetProAspArgGlyLysLys---ValLysIleGluSer 239
 4183 CTGCGGGCAGAGCTGCTCGGGGCCAGCGGAGCTGGGGAGCTGATCTCTCGGGCAG 4242
 240 TrpIleLysAspGlu-----AlaGluThrAlaAlaValAla 251
 4243 AAGGTGGCAGAGCAGGAGCAAGCAGCTCAGCAGCTGGCGGACAGAGAGCCAGCTATGCA 4302
 252 TrpArgLysArgLeuMetThrGluGly---GlyLeuAlaAlaGluLysMetAspAla 270
 4303 GAGCAGCTGAGCATGCTGAAGAGAGCGCCATGCTGTCGACAGAGAGAAC----- 4353
 271 ArgGlyLeuLeuLeuValAlaCysPheGlyValProSerAsnPheArgSerThrAsp 290
 4354 CGGGGCTG-----GGTCAGCGGCCCAACCTTGGCCGGCAGTTT 4392
 291 Leu-----LeuAspLeuIleArgMetSerGlySerAsnIleLeuAlaGlyAlaLeu 307
 4393 CTGGAAGTGGAGTTGGACCAAGCCCGGGAAGATATGTCGAAGAGTTGGCAGCCGTACGT 4452
 308 LysArgSerGlnPheLeuValProMetValSerGlyIleValGluSerIleLys--- 326
 4453 GCTGATGCTGAGACCGCTGCTGGCTGAGTGCAGCGAGAGCAGCAGACAGTCCCGGGAG 4512
 327 -----ArgGlyMetHisIleGluAlaLeuMetValTyr 338
 4513 CTGAGAGTGATGACTGCCAAGTATGAGGTGCCAAGGTCAAGGTCTGAGGAGAGGCGAG 4572
 339 ThrPheGlyMetGlu----- 343
 4573 CGGTTCAGGAGAGAGAGCAGAACTCACTGCCAGGTGGAGAACTGAGTAGAAGAACTG 4632
 344 ---AspLysPheSerAlaAlaLeuValLeuThrSerPheLeuLysMetSerLysGluSer 362
 4633 GCTGACTCTGACCAAGCCAGCAAGGTGTCAGCAGCAGAGAGCTGAAGGTCTCCAGGCTCAG 4692
 363 PheGluArgAlaLysArgLysAlaGlnSerProLeuAlaPheLysGluAlaAlaThrLys 382
 4693 GGAGCGAGAGCCAGGAGGCGCCAG-----CGCTTCAGGCGCCAGCTGAATGAA 4743
 383 GlnLeuAlaValLeuSerSerValMetGlnCysMetGluThrHisLysLeuAsp----- 400
 4744 CTGCAAGCCAGTTGAGCCAGAGGAGCAGCAGCTGAGCAGCTATAGCTGCAGATGGAG 4803
 401 -----ProAlaLysGluLeuProGlyTrpGlnIleLysGluGlnIle 414
 4804 AAAGCCAAACACATTATGATGCCAAGAGCAGGAGCAACCAAGAGAGTGCAGGAGCAGCTG 4863

415 ValSerLeuGluLysAspThrLeuGlnLeuAspLysGluMetGluGluLysAlaArgSer 434
 4864 CGGAGCTGGAG-----CAGCTGCAGAAAGAAACAAAGAG----- 4899
 435 LeuSerLeuMetGluGluAlaAlaLeuAla-LysArgMetTyrAsnGlnGlnIleLysAr 454
 4900 -----CTGCGAGCTGAAGCTGAACGGCTGGGCCATGAGCTACAGCAGGCTGGGCTGAAG 4953
 454 gProArgLeuSerProMetGluMetProProValThrSerSerSerTyrSerProIleTyr 474
 4954 ACCAAG-----AGGCTGAACAGACCTGCCGCCACCTTACTGCCCCAGGTGCGCAGCCTG 5007
 474 rArgAspArgSerPheProSerGlnArgAspAspGlnAspGluIleSerAla----- 492
 5008 GAGGCACAGS---TTGCCCATGCAGACAGCAGCTTCGAGAGCTGGGCAATTCAGGTG 5064
 493 -----LeuValSerSerTyrLeuGlyProSer----- 501
 5065 GCAACTGATGCTTTAAAGAGCGGTGAGCCCGCAGGCTAAGCCCGCAGCTGGACTTGAGTATT 5124
 502 -ThrSerPheProHisArgSerArgSerArgSerProGluTyrMetValPro----- 517
 5125 GACAGCTGTGATCTGAGCTCGGAGGAGGAGCCCACTCAGTATCACCAGCAGCTGCT 5184
 518 -LeuProHisGlyGlyLeuGlyArgSerValTyrAlaTyrGluHisLeuAlaProAsnSe 537
 5185 CGTACCCAGCAGCAGCGCACCGCTCCCTGGAGAACACAGCCTCACCCTATCTCCAGCGC 5244
 537 rTyrSerPro-GlyHisGlyHisArgLeuHisArgGlnTyrSerProSerLeuValHisG 557
 5245 CTGCCCCCAAGGT-----AGATCCCTGGAGAGTCTCTACTTCACT 5286
 557 LysGlnArgHisProLeuGlnTyrSerProIleHisGlyGlnGlnLeuProTyrG 577
 5287 CC-----CATCCCTGCTCGAGTTCAGGCCCTCGAGAGCAGCTGAGCTCCCTGGG- 5339
 577 LysLeuArgVal----- 581
 5340 -----AGACGCTCTCTGGAGCTCGGCTCGTAAGACCCGCTCGTGGCGCACACAG 5394
 582 -----TyrArgHisSerProSerGluGluArgTyrLeuGlyLeuSerAsnGlnArg 598
 5395 CAGTATCATCAATCATCAGCATGACCAAGAGAGTAGATGTGGAAGAGCCAGACAGCGC 5450

Search completed: August 25, 2003, 13:28:25
 Job time : 274 secs

ATTORNEY/AGENT INFORMATION:

NAME: PITCHER ESQ, EDMUND R
REGISTRATION NUMBER: 27,829
REFERENCE/DOCKET NUMBER: MTP-013
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/248-7000
TELEFAX: 617/248-7100
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 6306 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:

NAME/KEY: CDS
LOCATION: 1..6306
PUBLICATION INFORMATION:
AUTHORS: COMPTON, DUANE A
AUTHORS: SZILAK, ILLYA
AUTHORS: CLEVELAND, DON W
TITLE: PRIMARY STRUCTURE OF OF NIMA, AN INTRANUCLEAR
TITLE: PROTEIN THAT DEFINES A NOVEL PATHWAY FOR
TITLE: SEGREGATION OF PROTEINS AT MITOSIS
JOURNAL: JOURNAL OF CELL BIOLOGY
VOLUME: 116
PAGES: 1395-1408
DATE: MAR-1992
US-08-195-487-3

Alignment Scores:
Pred. No.: 0.059 Length: 6306
Score: 127.00 Matches: 161
Percent Similarity: 34.61% Conservative: 102
Best Local Similarity: 21.18% Mismatches: 275
Query Match: 4.03% Indels: 222
Gaps: 35

US-09-890-475-1 (1-609) x US-08-195-487-3 (1-6306)

2y 9 AlaAlaGlnProThrThrThrAlaAsnProLeuLeuGlnArgHisGlnSerglu----- 26
Db 3331 GCTGCTGCAGGACAGACGCCAACAGGCCCAAGCTGGAAGCACTGGCGGCAGAGGTGAGC 3390
2y 27 -----GlnArgArgGlu-----LeuProLysIleVal 36
Db 3391 AAGCTGGAACAGCAATGCCAGAGCAGCAGGAGCGCTGACAGCCTGGAACGACGCTC 3450
2y 37 GluThrGluSerThrSer-----MetAspIleThrIleGlyGln 49
Db 3451 GAGCTGAGCGGGCTCCCGGGCTGACGGGACAGTGTCTGTGAGACTCTGACGGGCCAG 3510
2y 50 -----SerLysGlnProGlnPheLeuLysSerIleAsp 60
Db 3511 TTAGAGGAGAGGCCAGGAGCTAGGCGACAGTCAGAGTGCCTTAGCTCGGCCAACGG 3570
2y 61 GluLeuAlaIlePheSerValAlaValGluThrPheLysArgGlnPheAspLeuGln 80
Db 3571 GAGTTGGTGTCTTCGGCACCAGGTTACAGACCACAGCAGGCTGAGATGAGTGGAG 3630
2y 81 LysHisIle-----GluSerIleGluAlaIleAspSerLysLeu 94
Db 3631 GCCCAGGTGGCCCGGGCCGCAAGAGGCTGAGAGGAAATAAGCTCATCAGCAGCTTG 3690
2y 95 GluSerAsnGlyValValLeuAlaAlaArgAsnAsnAsnPheHisGlnProMetLeuSer 114
Db 3691 GAGGAGGAGGTGTCATCCTG-----AATGCCAGGTCCTGGAG 3729
2y 115 ProProArgAsnAsnValSerValGluThrThrValThrVal-----SerglnProSergln 133
Db 3730 AAGGAGGGGAGACCAAGGAGTTGAAGCGGCTGTGATGTCGCGAGTCAGAGAAGCCAG 3789
2y 134 GluIle-----ValProGluThr 139

Db 3790 AAGCTGGAGGAGCTGCGCTGCTGTCAGGAGAGAGCAGCAGCAACAGTCCAGAGCTG 3849
Qy 140 SerAsnLysPro-----GluGlyGlyArgMetCys----- 149
Db 3850 CAGACACGAGCTCTGCTCTGCGGGAGGAGGTGCAGAGCTCCGGAGGAGGCTGAGAAA 3909
Qy 150 GluLeuMetCysSerLysGlyLeuArgLysValIleTyrAlaAsnIleSerAspGlnAla 169
Db 3910 CAGCGGTGGCTTCAGAGAACCTGCGGCAG-----GAGCTGACCTCAGAGCT 3957
Qy 170 LysLeuMetGluGluIleProSerAlaLeuLysLeuAlaLysGlu----- 184
Db 3958 GAGCGTGGAGGAGCTGGGCCAAGATTGAAGCGTGGCAGGAGAGTTCTTCCAGAAA 4017
Qy 184 ----- 184
Db 4018 GAGCAGGCCCTCTCCACCTGTCAGCTCGAGCACACACAGCCGCTGGTGGTGGAG 4077
Qy 185 -----ProAlaLysPheValLeuAspCysIleGlyLysPheTyrLeuGlnGlyArgArg 202
Db 4078 CTGCTGCCAGCTAAGCACCTCTGCCAG-----CAGCTGACGCCCGAGCAG 4122
Qy 203 AlaPheThrLysGluSerProMetSerSerAlaArgGlnValSerLeuLeuIle----- 220
Db 4123 GCGCTGCCGAGAAACGCCACCTGAGAGCTGAGAGCAGACAGCAGCGCTGGGGGA 4182
Qy 221 LeuGluSerPheLeuLeuMetProAspArgGlyLysGlyLys-----ValLysIleGluSer 239
Db 4183 CTGCGGCAGAGCTGCTGCGGGCCAGCGGAGCTGGGGAGCTGATTCCTCTCGCGCAG 4242
Qy 240 TrpIleLysAspGlu-----AlaGluThrAlaAlaValAla 251
Db 4243 AAGTGGCAGAGCAGACCGACAGCTCAGCAGCTGCGGGCAGAGCAGCAGCTATGCA 4302
Qy 252 TrpArgLysArgLeuMetThrGluGly-----GlyLeuAlaAlaGluLysMetAspAla 270
Db 4303 GAGCAGCTGAGCATGCTGAAGAAGCGCATGCTGCTGTCGAGAGGAGAAC----- 4353
Qy 271 ArgGlyLeuLeuLeuValAlaCysPheGlyValProSerAsnPheArgSerThrAsp 290
Db 4354 CGGGGGCTG-----GGTAGCGGGCCCACTTGGCGCGGAGTTT 4392
Qy 291 Leu-----LeuAspLeuIleArgMetSerGlySerAsnGluIleAlaGlyAlaLeu 307
Db 4393 CTGGAAGTGGAGTTGACCGAGCCCGGAAAGATATGCCAAGAGTTGGCAGCCCTAGCT 4452
Qy 308 LysArgSerGlnPheLeuValProMetValSerGlyIleValGluSerSerIleLys--- 326
Db 4453 GCTGATGCTGAGACCGCTCTGGGTGAGTGCAGCGAGAGCAGACAGCAGCACTGCCGGGAG 4512
Qy 327 -----ArgGlyMetHisIleGluAlaLeuGluMetValTyr 338
Db 4513 CTGAGGTGATGATCTGCCAAGTATGAGGTGCCAGGTCCAGGTCTGAGGAGGAGCAG 4572
Qy 339 ThrPheGlyMetGlu----- 343
Db 4573 CGGTTCCAGAGAGAGAGGAGCAAACTCACTGCCAGGTGGAAGAACTGAGTAAGAACTG 4632
Qy 344 ---AspLysPheSerAlaAlaLeuValLeuThrSerPheLeuLysMetSerLysGluSer 362
Db 4633 GCTGACTCTCCACCAAGCCAGAGGTGCGAGCAGCAGCAAGCTGAGGCTGTCAGGCTCAG 4692
Qy 363 PheGluArgAlaLysArgLysAlaGlnSerProLeuAlaPheLysGluAlaAlaThrLys 382
Db 4693 GGAGCGCAGAGCAGCAGGAGGAGGAGCAG-----CGTTCAGGCCCGCCGCTGAATGAA 4743
Qy 383 GlnLeuAlaValLeuSerSerValMetGlnCysMetGlnThrHisLysLeuAsp----- 400
Db 4744 CTGCAAGCCCGAGTTGAGCCAGAGAGGAGCGAGCTGAGCACTATAGCTGAGGATGGAG 4803
Qy 401 -----ProAlaLysGluLeuProGlyTrpGlnIleLysGluGlnIle 414

QY 140 SerAsnLysPro---GluGlyGlyArgMetCys----- 149
 Db 3850 CAGAACGACGCTCTGCTCTCGCGGAGAGAGTGCAGAGCCCTCCGGAGGAGGCTCAGAAA 3909
 QY 150 GluLeuMetCysSerLysGlyLeuArgLysTyrIleTyrAlaAsnIleSerAspGlnAla 169
 Db 3910 CAGCGGTGGCTTCAGAGAACCTTGGCCAG-----GAGTGACCTCAGAGCT 3957
 QY 170 LysLeuMetGluGlnIleProSerAlaLeuLysLeuAlaLysGlu----- 184
 Db 3958 GAGCGTCCGAGGAGCTGGGCCAAGAAATTGAAGCGTGGCAGAGAGAAATTCTTCAGAAA 4017
 QY 184 ----- 184
 Db 4018 GAGCAGGCCCTCTCCACCTCGAGCTCGAGCACACAGCAGCCCTGTGTAGTGAG 4077
 QY 185 -----ProAlaLysPheValLeuAspCysIleGlyPheTyrLeuGlnGlyArg 202
 Db 4078 CTGCTGCCAGCTAAGCACCTTGTCCAG-----CAGCTGAGGCCGAGCAG 4122
 QY 203 AlaPheThrLysGluSerProMetSerSerAlaArgGlnValSerLeuLeuIle----- 220
 Db 4123 GCCCTCCGAGAGAACGCCACCGTAGGAGCTGGAGCAGCAGCAGCCCGCTGGGGA 4182
 QY 221 LeuGluSerPheLeuLeuMetProAspArgGlyLysLys---ValLysIleGluSer 239
 Db 4183 CTGCGGCAGAGCTGCTGCGGCCACGCGGAGCTTGGGAGCTGATCTCTGCGGCAG 4242
 QY 240 TrpIleLysAspGlu-----AlaGluThrAlaAlaValAla 251
 Db 4243 AAGGTGGCAGCAGAGGAGCGAACAGCTCAGCAGCTCGCGGCAGAGAGCCAGCTATGCA 4302
 QY 252 TrpArgLysArgLeuMetThrGluGly---GlyLeuAlaAlaAlaGlyLysMetAspAla 270
 Db 4303 CAGCAGCTGACATGCTGCTGAAGAGCGCATGGCTGCTGCGCAGAGAAC----- 4353
 QY 271 ArgGlyLeuLeuLeuValAlaCysPheGlyValProSerAsnPheArgSerThrAsp 290
 Db 4354 CGGGGGCTG-----GGTGAGCGGCCAACCTTGGCCGCGGAGTTT 4392
 QY 291 Leu-----LeuAspLeuIleArgMetSerGlySerAsnGluIleAlaGlyAlaLeu 307
 Db 4393 CTGGAAGTGGAGTTGGACCGCGCGGAAAGATATGTCCAAGATTGGCAGCGTACGT 4452
 QY 308 LysArgSerGlnPheLeuValProMetValSerGlyIleValGluSerSerIleLys--- 326
 Db 4453 GCTGATGCTGAGACCGCTGCTGAGTGCAGCGCAGCAGCAGCAGCACTGCGCGGAG 4512
 QY 327 -----ArgGlyMetHisIleGluAlaLeuGluMetValTyr 338
 Db 4513 CTGGAGGTGATGATCGCCAAAGTATGAGGTGCCAAGTCAAGTCTCGGAGGAGGCGAG 4572
 QY 339 ThrPheGlyMetGlu----- 343
 Db 4573 CGGTTCCAGGAAGAGGAGCAGAACTACTCTGCCAGGTGGAAGAACTAGTAAGAAACTG 4632
 QY 344 ---AspLysPheSerAlaAlaLeuValLeuThrSerPheLeuLysMetSerLysGluSer 362
 Db 4633 GCTGACTCTGACCAAGCAGCAAGGTGCAGCAGAGAGCTGAAGGCTGTCCAGGCTCAG 4692
 QY 363 PheGluArgAlaLysArgLysAlaGlnSerProLeuAlaPheLysGluAlaAlaThrLys 382
 Db 4693 GGAGCGGAGAGCAGCAGAGGCCAG-----CGCTCCAGGCCAGCTGAATGAA 4743
 QY 383 GlnLeuAlaValLeuSerSerValMetGlnCysMetGluThrHisLysLeuAsp----- 400
 Db 4744 CTGCAAGCCCACTTGAAGCCAGAGCAGGAGCGAGCTGAGCACTATAAGCTGCAGATGGAG 4803
 QY 401 -----ProAlaLysGluLeuProGlyTyrGlnIleLysGluGlnIle 414
 Db 4804 AAAGCCAAACATATGATGCCAAGAGCAGCAGAGCAGCAGAGCTGCAGGAGCAGCTG 4863

QY 415 ValSerLeuGluLysAspThrLeuGlnLeuAspLysGluMetGluGluLysAlaArgSer 434
 Db 4864 CGGAGCTGGAG-----CAGCTGCAGAGGAAACAAAGAG----- 4899
 QY 435 LeuSerLeuMetGluGluAlaAlaLeuAla-LysArgMetTyrAsnGlnGlnIleLysArg 454
 Db 4900 -----CTGCGAGCTGAAGCTGAACGGCTGGGCCCATGATGACAGCAGGCTGGGCTGAAG 4953
 QY 454 gProArgLeuSerProMetGluMetProValThrSerSerSerTyrSerProIleTyr 474
 Db 4954 ACCAAG-----AGCTGAACAGACCTGCCCGCACCTTACTGCCAGGTGCCGAGCTG 5007
 QY 474 rArgAspArgSerPheProSerGlnArgAspAspGlnAspGluLysSerAla----- 492
 Db 5008 GAGGCACAGG---TTGCCCATGACAGCAGCAGCTTCGAGACCTGGGCAATTCAGGTG 5064
 QY 493 -----LeuValSerSerTyrLeuGlyProSer----- 501
 Db 5065 GMACTGATGCTTTAAGAGCCGTGAGCCCGCAGGCTAAGCCCGAGCTGGACTTGAGTATT 5124
 QY 502 -ThrSerPheProHisArgSerArgSerProGluTyrMetValPro----- 517
 Db 5125 GACAGCTGGATCTGAGCTGCGAGGAGGGGACCCCACTCAGTATCCACGACGAAGCTGCCT 5184
 QY 518 -LeuProHisGlyGlyLeuGlyArgSerValTyrAlaTyrGluHisLeuAlaProAsnSe 537
 Db 5185 CGTACCCAGCCAGACCGCAGCGCTCCCTGGAGAACCCAGCTCACCCTATCTCCAGGCG 5244
 QY 537 rTyrSerPro-GlyHisGlyHisArgLeuHisArgGlnTyrSerProSerLeuValHisG 557
 Db 5245 CTGCCCCCAAGGT-----AGAATCCCTGGAGAGTCTCTACTTCACT 5286
 QY 557 LysGlnArgHisProLeuGlnTyrSerProIleHisGlyGlnGlnLeuProTyrG 577
 Db 5287 CC-----CATCCCTCTCGAGTCCAGGCCCCCCTGGAGAGCAGCTCGACTCCCTGGG- 5339
 QY 577 LysLeuArgVal----- 581
 Db 5340 -----AGAGCTCTCTCGAGCTCGGTCTGTAAGACCGCTCCGCTCGTGGCGCACCAG 5394
 QY 582 ---TyrArgHisSerProSerGluGluArgTyrLeuGlyLeuSerAsnGlnArg 598
 Db 5395 CAGATCATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAAT 5450
 RESULT 14
 US-08-195-487-3
 ; Sequence 3, Application US/08195487
 ; Patent No. 5783403
 ; GENERAL INFORMATION:
 ; APPLICANT: TOKUATLY, GARY
 ; APPLICANT: LIDGARD, GRAHAM P
 ; TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE
 ; TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX
 ; NUMBER OF SEQUENCES: 6
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: TESTA HURWITZ & THIBEAULT
 ; STREET: 53 STATE STREET
 ; CITY: BOSTON
 ; STATE: MA
 ; COUNTRY: USA
 ; ZIP: 02109
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent in Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/195,487
 ; FILING DATE:
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US/07/901,701
 ; FILING DATE:

REGISTRATION NUMBER: 27,829
REFERENCE/DOCKET NUMBER: MTP-013
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 248-7000
TELEFAX: (617) 248-7100
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 6306 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:

NAME/KEY: CDS
LOCATION: 1..6306
PUBLICATION INFORMATION:
AUTHORS: COMPTON, DUANE A
AUTHORS: SZILAK, ILYA
AUTHORS: CLEVELAND, DON W
TITLE: PRIMARY STRUCTURE OF NUMA, AN INTRANUCLEAR
TITLE: PROTEIN THAT DEFINES A NOVEL PATHWAY FOR
TITLE: SEGREGATION OF PROTEINS AT MITOSIS
JOURNAL: J. Cell Biol.
VOLUME: 116
PAGES: 1395-1408
DATE: 1992
US-08-467-781-3

Alignment Scores:
Pred. No.: 0.059
Score: 127.00
Percent Similarity: 34.61%
Best Local Similarity: 21.18%
Query Match: 4.09%
DB: 1
Length: 6306
Matches: 161
Conservative: 102
Mismatch: 225
Indels: 222
Gaps: 35

US-09-890-475-1 (1-609) x US-08-467-781-3 (1-6306)

QY 9 AlaAlaGlnProThrThrThrAlaAsnProLeuLeuGlnArgHisGlnSerGlu----- 26
Db 3331 GCTGCTGGCAGGACAGAGCCCAACAGCGCCCAAGCTGGAAGCAGCTGCGGCAGGTGAGC 3390
QY 27 -----GlnArgArgArgGlu-----LeuProLysIleVal 36
Db 3391 AAGTGTGAACAGCAATCCCAAGAGCAGCAGCAGCTGACAGCTTGGACGCCTC 3450
QY 37 GluThrGluSerThrSer-----MetAspIleThrIleGln 49
Db 3451 CAGGCTGAGCGGCGCTCCCGGCTGAGCGGACAGTCTCTGGAGACTCTGAGGCGCCAG 3510
QY 50 -----SerLysGlnProGlnPheLeuLysSerIleAsp 60
Db 3511 TTAGAGGAGAGAGCCAGAGCTAGGCGCAGTCAGAGTCCCTTACCTCGGCCCAACGG 3570
QY 61 GluLeuAlaAlaPheSerValAlaValGluThrPheLysArgGlnPheAspLeuGln 80
Db 3571 GAGTTGGCTGCTTCCGACCAAGGTACAGACACAGCAAGCTGAAGATGATGAGTGAAG 3630
QY 81 LysHisIle-----GluSerIleGluAsnAlaIleAspSerLysLeu 94
Db 3631 GCCCAGGTGGCGCGCGGCGGCGGCTGAGAGGAAATAGCTCTATCAGCAGCTTG 3690
QY 95 GluSerAsnGlyValValLeuAlaAlaArgAsnAsnPheHisGlnProMetLeuSer 114
Db 3691 GAGGAGGAGGTGTCCATCTCTG-----AATCGCCAGGTCTCTGGAG 3729
QY 115 ProProArgAsnAsnValSerValGluThrThrValThrVal-----SerGlnProSerGln 133
Db 3730 AAGGAGGCGGAGAGCAGAGGAGTGAAGCGGCTGTGTGATGGCCGAGTCCAGAGAGCCAG 3789
QY 134 GluIle-----ValProGluThr 139
Db 3790 AAGCTGGAGGAGAGCTGCGCTGTGTGAGGAGAGAGCAGCCAGCACAGTCCAGAGCTG 3849

415 ValSerLeuGluLysAspThrLeuGlnLeuAspLysGluMetGluGluLysAlaArgSer 434
4864 CGGAGCTTGAG-----CAGCTCAGAGGAAACAAGAG----- 4899
435 LeuSerLeuMetGluGluAlaAlaLeuAla-LysArgMetTyrAsnGlnGluIleLysArg 454
4900 -----CTGGCAGCTGAAGCTGAACGCTGGCCATCATGAGCTACAGCAGCGTGGCTGAAG 4953
454 gProArgLeuSerProMetGluMetProProValThrSerSerSerTyrSerProIleTyr 474
4954 ACCAAG-----AGGTGACACAGCTGCGCCACCTTACTGCCAGGTGGCGACCTG 5007
474 rArgAspArgSerPheProSerGlnArgAspAspGlnAspGluIleSerAla----- 492
5008 GAGGCACAGG-----TTGCCCATGACAGCAGCAGCAGCTTCGAGACCTGGGCAAAATTCAGGTG 5064
493 -----LeuValSerSerTyrLeuGlyProSer----- 501
5065 GCAACTGATGCTTTAAGAGCGGTGAGCCCGCAGGCTAAGCCCGAGCTGGACTTGAGTATT 5124
502 -ThrSerPheProHisArgSerArgArgSerProGluTyrMetValPro----- 517
5125 GACAGCTGGATCTGAGTGGCAGGAGGAGGAGCCCACTCAGTATCACCAGCAAGCTGCGCT 5184
518 -LeuProHisGlyGlyLeuGlyArgSerValTyrAlaTyrGluHisLeuAlaProAsnSe 537
5185 CGTACCCAGCAGCAGCAGCAGCAGCGTCCCTGGAGAACCACTCCTATCTCCAGCGC 5244
537 rTyrSerPro-GlyHisGlyHisArgLeuHisArgGlnTyrSerProSerLeuValHisG 557
5245 CTGCCCCCAAGGT-----AGAATCCCTGGAGAGTCTCTACTTCACT 5286
557 lyGlnArgHisProLeuGlnTyrSerProProIleHisGlyGlnGlnGlnLeuProTyrG 577
5287 CC-----CATCCCTGCTCGGAGTCAGGCGCCCTGGAGAGCAGCTGGACTCCCTGGG- 5339
577 lyIleGlnArgVal----- 581
5340 -----AGAGCTCTCTCGACTCGGTGTAAGACCGCTCGCTCGTGGCGCACCAGC 5394
582 -----TyrArgHisSerProSerGluGluArgTyrLeuGlyLeuSerAsnGlnArg 598
5395 CAGATCATCAATCATCATCAATGACCAAGAGCTAGATGTGGAAGAGCCAGCAGCGC 5450

RESULT 13
US-08-467-781-3
Sequence 3, Application US/08467781
Patent No. 5780596
GENERAL INFORMATION:
APPLICANT: TOUTKATLY, GARY
APPLICANT: LIDGARD, GRAHAM P
TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE
TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSES: TESTA, HURWITZ & THIBEAULT
STREET: 125 HIGH STREET
CITY: BOSTON
STATE: MA
COUNTRY: USA
ZIP: 02110
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/467,781
FILING DATE: 06-JUN-1995
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: PITCHER ESQ, EDMUND R

NAME: PITCHER ESQ, EDMUND R
 REGISTRATION NUMBER: 27,829
 REFERENCE/DOCKET NUMBER: MTP-013
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (617) 248-7000
 TELEFAX: (617) 248-7100
 INFORMATION FOR SEQ ID NO: 3:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 6306 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 1..6306
 PUBLICATION INFORMATION:
 AUTHORS: COMPTON, DUANE A
 AUTHORS: SZILAK, ILYA
 AUTHORS: CLEVELAND, DON W
 TITLE: PRIMARY STRUCTURE OF NUMA, AN INTRANUCLEAR
 TITLE: PROTEIN THAT DEFINES A NOVEL PATHWAY FOR
 TITLE: SEGREGATION OF PROTEINS AT MITOSIS
 JOURNAL: J. Cell Biol.
 VOLUME: 116
 PAGES: 1385-1408
 DATE: 1992
 US-08-470-950-3

Alignment Scores:

Pred. No.: 0.059 Length: 6306
 Score: 127.00 Matches: 161
 Percent Similarity: 34.61% Conservative: 102
 Best Local Similarity: 21.18% Mismatches: 275
 Query Match: 4.09% Indels: 222
 Gaps: 35

US-09-890-475-1 (1-609) x US-08-470-950-3 (1-6306)

QY 9 AlaAlaGlnProThrThrAlaAsnProLeuLeuGlnArgHisGlnSerGlu----- 26
 DB 3331 C T C T G C G C A G C A G A G C C A A C A G C C C C A G C T G G A G C A C T C G G G C A G A G T G A G C 3390
 QY 27 -----GlnArgArgGlu-----LeuProLysIleVal 36
 DB 3391 A A G C T G G A A C A G C A A T C C C A G A A G C A G C A G C A G C A G C T G A C A G C C T G G A A C G A G C C T C 3450
 QY 37 GluThrGluSerThr-----MetAspIleThrIleGlyGln 49
 DB 3451 G A G C T G A G C G G G C T C C C G G G C T G A G C G G C A G C A G T G C T C T G A G A C T C T G A G C C C A G 3510
 QY 50 -----SerLysGlnProGlnPheLeuLysSerIleAsp 60
 DB 3511 T T A G A G A G A A G C C C A G A G C T A G G C C A C A G T C A G A G T G C C T T A G C C T C G G C C C A A C G G 3570
 QY 61 GluLeuAlaAlaPheSerValAlaValGluThrPheLysArgGlnPheAspAspLeuGln 80
 DB 3571 G A G T G G C T G C T T C C C A C C A A G T A C A G C A C C A G C A G C T G A A G A T C A G T G G A A G 3650
 QY 81 LysHisIle-----GluSerIleGluAsnAlaIleAspSerLysLeu 94
 DB 3631 G C C A G G T G C C C G G C C G C C A A G A G G T G A G A G A A A A T A G C C T A T C A G C A G C T T G 3690
 QY 95 GluSerAsnGlyValValLeuAlaAlaArgAsnAsnAsnPheHisGlnProMetLeuSer 114
 DB 3691 G A G G A G A G G T G T C C A T C T G -----AATCGCCAGGTCTCTGGAG 3729
 QY 115 ProProArgAsnAnValSerValGluThrThrValThrVal-----SerGlnProSerGln 133
 DB 3730 A A G A G G G G A G A C A A G A G A G T T A A A G C G G C T G T A T G G C C G C A G A G A G C C A G 3789
 QY 134 GluIle-----ValProGluThr 139
 DB -----ValProGluThr 139

DB 3790 A A G C T G A G A G A G A G C T C G C C T C T C G A G C G C A G A C A C C A G C A A C A G T G C C A G A G C T G 3849
 QY 140 SerAsnLysPro---GluGlyGlyArgMetCys----- 149
 DB 3850 C A G A A C G C A G C T C T C T C T C G C G A G A G T G C A G A C C T C C G G A G G A G G C T C A G A A A 3909
 QY 150 GluLeuMetCysSerLysGlyLeuArgLysTyrIleTyrAlaAsnIleSerAspGlnAla 169
 DB 3910 C A G C G G T G G C T T C A G A A C C T G C G G C A G -----GAGCTGACCTCAGAGCT 3957
 QY 170 LysLeuMetGluGluIleProSerAlaLeuLysLeuAlaLysGlu----- 184
 DB 3958 G A G C G T C G G A G A G C T G G C C A A G A A T T G A A G C G C T G C A G A G A A G T T C T T C C A G A A 4017
 QY 184 ----- 184
 DB 4018 G A G C A G C C C T C T C C A C C C T G C A G C T G C A G C A C C A C A G C A G C C C T G T G A G T A G A G 4077
 QY 185 -----ProAlaLysPheValLeuAspCysIleGlyLysPheTyrIleuGlnGlyArg 202
 DB 4078 C T G T G C C A G C T A A G C A C C T C T G C C A G -----CAGCTGACGCCGAGCAG 4122
 QY 203 AlaPheThrLysGluSerProMetSerSerAlaArgGlnValSerLeuLeuIle----- 220
 DB 4123 C C C G C T C C C A G A A A C C C A C C G T G A G A G C T G C A G C A G C A A C A G C A G C C G C T G G G G A 4182
 QY 221 LeuGluSerPheLeuLeuMetProAspArgGlyLysGlyLys-----ValLysIleGluSer 239
 DB 4183 C T G C G G C A G A G C T G C T G C G G C C C A G C G A G C T T G G G A G C T G A T T C C T C T C G G C A G 4242
 QY 240 TrpIleLysAspGlu-----AlaGluThrAlaAlaValAla 251
 DB 4243 A A G T G G C A G C A G A G C A A C A G C T C A G C A G C T G C G G C A G A G A G C C A G C A T A T G C A 4302
 QY 252 TrpArgLysArgLeuMetThrGluGly---GlyLeuAlaAlaAlaGluLysMetAspAla 270
 DB 4303 G A G C A G C T G A G C A T G T G A A A G A G C G C A T G C C T G C T G C A G A G A A A C ----- 4353
 QY 271 ArgGlyLeuLeuLeuValAlaCysPheGlyValProSerAsnPheArgSerThrAsp 290
 DB 4354 C G G G G C T G -----G G T A G C G G G C C A C C T T G C C G C G C A G T T 4392
 QY 291 Leu-----LeuAspLeuIleArgMetSerGlySerAsnGluIleAlaGlyAlaLeu 307
 DB 4393 C T G A A G T G G A G T T G G A C C A G C C C G G A A A A G T A T G T C C A A G A G T T G C A G C C C T A C G T 4452
 QY 308 LysArgSerGlnPheLeuValProMetValSerGlyIleValGluSerSerIleLys--- 326
 DB 4453 G C T G A T C T G A G A C C C T C T G G C T G A G T G C A G C A G A G A C A G A G C A C T C C C G G A G 4512
 QY 327 -----ArgGlyMetHisIleGluAlaLeuGluMetValTyr 338
 DB 4513 C T G A G G T G A T A G C T G C C A A G T A T A G G G T G C C A A G T C A A G G T C C T G A G A G A G G C A G 4572
 QY 339 ThrPheGlyMetGlu----- 343
 DB 4573 C G G T T C C A G A A G A G A G C C A G A A A C T A C T C C C A G G T G G A A A C T A G T A A G A A A C T G 4632
 QY 344 ---AspLysPheSerAlaAlaLeuValLeuThrSerPheLeuLysMetSerLysGluSer 362
 DB 4633 G C T G A C T G T C A C C A A G C C A G A G G T G C A G C A G A G A G C T G T C C A G C C T C A G 4692
 QY 363 PheGluArgAlaLysArgLysAlaGlnSerProLeuAlaPheLysGluAlaAlaThrLys 382
 DB 4693 G A G C G A G A G C C A G C A G A G G C C C A G -----C G T T C C A G C C C A G C T G A T A A G A A 4743
 QY 383 GlnLeuAlaValLeuSerSerValMetGlnCysMetGluThrHisLysLeuAsp----- 400
 DB 4744 C T G A A G C C C A G T T G A C C A G A A G A G C A G C A G C T G A C A C T A T A A G C T G C A G A T G A G 4803
 QY 401 -----ProAlaLysGluLeuProGlyTrpGlnIleLysGluGlnIle 414
 DB 4804 A A G C C A A A C A C A T T A T G A T G C C A A G A A G C A G A C A A C A A G A G C T G C A G A G C A G C T G 4863

Db 2877 CGAACACTTCTGTACAATACTTACGAAGGTATCGATTGTCGGGACTCTTCCGAAAG 2936
QY 394 MetGluThrHisLeuAspProAlaLysGluLeuProGlyTrpGlnIleLysGluGln 413
Db 2937 CTGGAGCTGCAGCTGCTGAGTGGAGCGTGCAGCGGATGCTCGGAGTGAAGGAGCGT 2996
QY 414 lleValSerLeu-----GluLysAspThrLeuGlnLeu 424
Db 2997 GTTCTGTCCAATGCCGTGCCAGCTAGCTAGTGAAGCGCAAGGAATCGGTGCTG 3056
QY 425 AspLysGluMetGluGluLysAlaArgSerLeuSerLeuMet-----Glu 439
Db 3057 CAGCGCCAGCAATGAA-----CTCATTGTGCTGCTGCATCGCGGAG 3107
QY 440 GluAlaLeuAlaLysArgMetTyrSerGlnGlnIleLysArgProArgLeuSerPro 459
Db 3108 GAAAGCTCGCCAGTTCGCCACCTACGAGACGAGCTTTGGCCCT-----GCC 3158
QY 460 MetGluMetProProValThrSerSer-----TyrSerProIleTyrArgAsp 476
Db 3159 TGTGTGCCCCCGGTGCAGCCAGTAGTGTGACTAGCATCTTCTACAAGGT 3218
QY 477 ArgSerPheProSerGlnArgAspAspGlnAspGluIleSerAlaLeuValSerSer 496
Db 3219 GCGCGCAATCCAAAGCAAG-----GCACGTATC----- 3245
QY 497 TyrLeuGlyProSerThrSerPheProHisArgSerArgArgSerProGluTyrMetVal 516
Db 3246 -----AAAGGCCAAG----- 3257
QY 517 ProLeuProHisGlyGlyLeu-----GlyArg 525
Db 3258 CCGATCCACCGCTCGCGCTAGACAATAACAGACAAAGAGCAACGCGGACGC 3317
QY 526 SerValTyrAlaTyrGluHisLeu-----AlaProAsnSerTyrSer 539
Db 3318 AACATTTGCCAAGTTCATGTGCTCACCAGATATGCTGCGGAGCGGTTC 3377
QY 540 ProGly--HisGlyHisArgLeuHisArgGlnTyrSerProSerLeuVal-----HisG 557
Db 3378 TCAGGTCCAGCGCGCAGGAATCCAC-----GGCCCTGATTGATCAACCACCACG 3428
QY 557 lylClnArgHisProLeuGlnTyrSerProProIle 568
Db 3429 GCCA-CGGCAACCACTACGTGCACACACCACCACTG 3462

RESULT 11

US-08-466-390-3
; Sequence 3, Application US/08466390
; Patent No. 568562
; GENERAL INFORMATION:
; APPLICANT: TOKUATLY, GARY
; TITLE OF INVENTION: LIDGARD, GRAHAM P
; TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE
; NUMBER OF INVENTION: INTERIOR NUCLEAR MATRIX
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TESTA, HURWITZ & THIBEAULT
; STREET: 125 HIGH STREET
; CITY: BOSTON
; STATE: MA
; COUNTRY: USA
; ZIP: 02110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/466,390
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:
NAME: PITCHER ESQ. EDMUND R
REGISTRATION NUMBER: 27,829
REFERENCE/DOCKET NUMBER: MTP-013
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 248-7000
TELEFAX: (617) 248-7100
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 6306 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..6306
PUBLICATION INFORMATION:
AUTHORS: COMPTON, DUANE A
AUTHORS: SZILAK, ILYA
AUTHORS: CLEVELAND, DON W
TITLE: PRIMARY STRUCTURE OF NUMA, AN INTRANUCLEAR
TITLE: PROTEIN THAT DEFINES A NOVEL PATHWAY FOR
TITLE: SEGREGATION OF PROTEINS AT MITOSIS
JOURNAL: J. Cell Biol.
VOLUME: 116
PAGES: 1395-1408
DATE: 1992
US-08-466-390-3
Alignment Scores:
Pred. No.: 0.059 Length: 6306
Score: 127.00 Matches: 161
Percent Similarity: 34.61% Conservative: 102
Best Local Similarity: 21.18% Mismatches: 275
Query Match: 4.09% Indels: 222
DB: 1 Gaps: 35
US-09-890-475-1 (1-609) x US-08-466-390-3 (1-6306)
QY 9 AlaAlaGlnProThrThrAlaAsnProLeuGlnArgHisGlnSerGlu----- 26
Db 3331 GCTGCTGCGAGCAGACAGACCCCAAGCTGGAAGCACTGCGGGCAGAGGTGAGC 3390
QY 27 -----GlnArgArgGlu-----LeuProLysIleVal 36
Db 3391 AGCTGGAACAGCAATGCCAAGCAGCAGCAGCTGACAGCTGACAGCTGACAGCCTC 3450
QY 37 GluThrGluSerThrSer-----MetAspIleThrIleGlyGln 49
Db 3451 GAGGCTGAGCGGGCTCCCGGGCTGAGCGGACAGTCTCTGGAGACTCTGCGAGGCCAG 3510
QY 50 -----SerLysGlnProGlnPheLeuLysSerIleAsp 60
Db 3511 TTAGAGGAGAGCCCGCAGAGCTAGGCGACAGTCAAGTGCCTTAGCTCGGCCAACGG 3570
QY 61 GluLeuAlaAlaPheSerValAlaValGluThrPheLysArgGlnPheAspLeuGln 80
Db 3571 GAGTTGGCTGCTTCCGCAACCAAGCTACAGCACCAAGAGCTGAAGATGATGAGGAAG 3630
QY 81 LysHisIle-----GluSerIleGluAsnAlaIleAspSerLysLeu 94
Db 3631 GCCCAGGTGGCCCGCGGCGGCGCAGAGCTGAGAGGAAATAATAGCCCTCATCAGCAGCTT 3690
QY 95 GluSerAsnGlyValValIleAlaAlaArgAsnAsnAsnPheHisGlnProMetLeuSer 114
Db 3691 GAGGAGGAGGTGCTCATCTG-----AATCGCAGGCTCTGGAG 3729
QY 115 ProProArgAsnAsnValSerValGluThrThrValThrVal-----SerGlnProSerGln 133
Db 3730 AAGGAGGGGAGAGCAGCAAGAGTGAAGCGGTGGTGTATGCGCGGATCAGAGAGCCAG 3789
QY 134 GluIle-----ValProGluThr 139

```

1260 ATCAAGTTCAGAAATCCAGCCCTGCTACATCTCTCTTTGTAGAACTTGATGAAGAAG 1319
410 Ilelys-----GluGlnIleValSerLeuGluLysasp 420
1320 GTGAAGACGACCAACTGAACCTACTGTAAGTGGAGGAGGTGTCTCCCGCACTGGAACCC 1379
421 ThrLeuGlnLeuAspLysGluMetGluGluLysAlaArgSerLeuMetGluGlu 440
1380 ACAGAAACGCTGCTGAGTGAAGAGGAGATAAACGCAAGGAGAGAGCCTTGTGAAGAG 1439
441 AlaAlaLeuAlaLysArgMetTyAsnGlnGlnIleLysArgPro----- 455
1440 CTGTCCTCCCTGCCACGAG-----AAGAAGCCCGTGGCCGCTGTGAG 1481
456 -----ArgLeuSerProMet--GluMetProValThrSerSer 468
1482 GGCAAGTCTAGACTGTCCCGCGGTGAGTGAAGCCCATGCCGCTGTCT 1532
RESULT 10.
JS-09-098-901-1
Sequence 1, Application US/09098901B
Patent No. 6218144
GENERAL INFORMATION:
APPLICANT: Scott, Matthew
APPLICANT: Sisson, John C.
TITLE OF INVENTION: Costal2 Genes and their Uses
FILE REFERENCE: SUN-65P
CURRENT APPLICATION NUMBER: US/09/098,901B
EARLIER APPLICATION NUMBER: 1998-06-17
EARLIER FILING DATE: 1997-06-30
NUMBER OF SEQ ID NOS: 11
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 1
LENGTH: 3605
TYPE: DNA
ORGANISM: D. Melanogaster
JS-09-098-901-1

Alignment Scores:
Pred. No.: 0.0207 Length: 3605
Score: 127.50 Matches: 150
Percent Similarity: 35.48% Conservative: 103
Best Local Similarity: 21.04% Mismatches: 244
Query Match: 4.11% Indels: 218
DB: 3 Gaps: 34

JS-09-890-475-1 (1-609) x US-09-098-901-1 (1-3605)
2Y 18 ProLeuLeuGlnArgHisGlnSerGluGlnArgArgGluLeuProLysIleValGlu 37
1492 CCCATTCTGGAGCGCGGAGGAGCCCGAGACGCTGAATCTCGAGAGCAATTCGGAGTCG 1551
38 ThrGluSerThrSerMetAspIleThrIleGlnSerLysGlnProGlnPheLeuLys 57
1552 CCAAACTCCGACACGAAACGACACACACAAATGAGTCGATCGGCCGCTCGGACGAC 1611
58 SerIleAspGluLeuAlaAlaPheSerValAlaValGluThrPheLysArgGlnPheAsp 77
1612 AAGATAGAAAGTCTA-----ATGGAAGAGTTTCGCGACAAAACAGAC 1653
78 Asp-----LeuGlnLysHisIleGluSerIleGluAsn 88
1654 GCTCTTATCTGTAAGAAACACGCTGAATATCTATCCAAAGCATCCGAGCGGCTTATGCAA 1713
89 AlaIleAspSerLysLeuGluSer-----ArgLys-----Val 99
1714 AGCCAAAGACCGGAGATTGGGCGACGCCCGCAGCAAGAAATGTTGATGATCGAAAGTC 1773
100 ValLeuAlaAlaArgAsnAsnAsnPheHisGlnProMetLeuSerProProArgAsnAsn 119
1774 AGCATTGGCAGTCGCGAGGAGAGTGT---CAGCCAGGTGCTGCTTAAGTACTGCTGAG 1830

```

```

120 ValSerValGluThrThrValThrValSerGlnProSerGlnGluIleValProGluThr 139
1831 CTTCGCATGCTTAATCGGGTAGCTTCCAGCAGCCGCTCCGCCCATGATCCTGAGTCG 1890
140 -----SerAsnLysProGluGlyArgMetCysGluLeuMetCys 153
1891 GTCGTCGATCCTCTGGAAGATTTCTTCGGCGGAGGAATCCGTCAGGCGGCTCTCGCTGCC 1950
154 Ser-----LysGlyLeuArgLysTyIleTyAla 163
1951 GCCCGCCGACCTCTCTATTGAACAGCTGCAGAAAAAATTCGCGCAACTGGCT---GCT 2007
164 AsnIleSerAspGlnAlaLysLeuMetGluGluIle----- 175
2008 GAGATCGAGGCAACACACAGCAGTTACGGGAAATCGAAGAACATCCAGGTAAACAAA 2067
176 -----ProSerAlaLeuLys-----Leu 181
2068 ATATAATCGCCGAATTG-GTCAAGAACACGCGATACAGCAGCATGCAAGCAAAAGATT 2126
182 AlalysGluProAlalysPheValLeuAspCys-----IleGlyLys 195
2127 CACAGAAACGTCGCAAACTTGAGCGCGAGTGCACAAAGCCCAAGAGCAGTTAGTAG 2186
196 PheTyLeuGlnGlyArg----- 201
2187 GCCTAGTTCAAGCGCGGGTCACTCGGAGATTGAGCGATGAGCACCACCAATATCGGACAT 2246
202 -----ArgAlaPheThrLysGluSerProMet-----SerSerAlaArg 214
2247 CTCGAGCGTCGACTAGAACCTCAGCTCAATGAAGCATATTGCGGTGAGAGCGGACAG 2306
215 GlnValSerLeuLeu-----IleLeuGluSerPheLeuLeuMetProAspArgGly 231
2307 AAGGTGAAGAGCTACAGCAATCGTGGCGGAGTCCGAAACACAGCCCGATATTACAG 2366
232 LysGlyLysValLysIleGluSerTrpIleLysAspGluAlaGluThrAlaAlaValAla 251
2367 AAA---AAGTTCGAAGAGGTGCAAGCTGCGCTGCCATGAGGCGGCGGTGTCATA 2423
252 TrpArgLysArgLeuMetThrGluGlyLeuAlaAlaGlu-----LysMet 268
2424 CTACGAGAACCCAGGAGACTGGCAAGAGAGTAGTAGAGCGCAAGGTTCTCCCGAGCAA 2483
269 AspAlaArgGlyLeuLeuLeuValAlaCysPheGlyValProSerAsnPheArgSer 288
2484 CAGGCGCCGAGTTAAAGCGAGTACAGCT-----AGGATT 2519
289 ThrAspLeuLeuAspLeuIleArgMetSerGlySerAsn---GluIleAlaGlyAlaLeu 307
2520 ACGCACCTTAATCACATTTTACCGGAGAGTTCGATAACCTGAGRAGCAGCGCGGACCA 2579
308 LysArgSerGlnPheLeu-----ValProMetValSerGlyIleValGluSerSer 324
2580 GAACGAGGAGACCTTGTGCTGATGATGATCGGCAACTTCGGCGGAACTCGTACTTTGTTG 2639
325 IleLysArgGlyMetHis----- 330
2640 TTGAAGACGCTGCTCATTTGGACCGCAAACTTAAGCGGACAAAGGTGCTGACGCCAAAAG 2699
331 -----IleGluAlaLeuGluMetValTyThr 339
2700 GAGGAGCGCAAGCTGCTCGAGTCGATGAGCCATGAGCCATGATAGTCGGCCATAGAA 2759
340 PheGlyMetGluAspLysPheSerAlaAlaLeuValLeuThrSerPheLeuLysMetSer 359
2760 TTCAAGACGAGATGATCAGCGGCCCGCTCCATCGACACGAGCGAC---CGAAATTCAG 2816
360 LysGluSerPheGluArgAlaLysArgLysAlaGlnSerProLeuAlaPheLysGluAla 379
2817 CGGAGAGGAGGAGAACAGATGCTGATGCGCACGCGCTAAATGCTCTCTCAACGGAGGAGTG 2876
380 AlaThrLysGlnLeuAlaValLeuSerSerValMetGlnCys----- 393

```

Qy 555 ValHisGlyGlnArgHisProLeuGlnTyrSerProProlleHisGlyGlnGlnGlnLeu 574
Db 2524 GTGATGATGCAAGCACTACCCAGGTTCTTCTACCTGCCTTCCAGCTACTCTTTTTC-----2577
Qy 575 ProTyrGlyIleGln-----Arg 580
Db 2578 CCAATATGGCAGCAGGTCTCAGGTGGTGAAGATGCTGACAGGCACGAGCCAGCCCGCAGT 2637
Qy 581 ValTyrArgHisSerProSerGluGluArgTyrLeuGlyLeuSerAsnGlnArgSerPro 600
Db 2638 GTGACTTGTAAATCCAGCTCAGAGTCCAAAGCCCTGGACATCTTGCAGCAGCAGTCCCGT 2697
Qy 601 ArgSerAsnSer 604
Db 2698 CACTACAGAGGC 2709

RESULT 9
US-09-620-312D-1020
; Sequence 1020, Application US/09620312D
; Patent No. 6569662
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyan
; APPLICANT: Chen, Rui-hong
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wehrman, Tom
; APPLICANT: Xue, Aidong J.
; APPLICANT: Yang, Yonghong
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Zhou, Ping
; APPLICANT: Ma, Yungqing
; APPLICANT: Wang, Dunrui
; APPLICANT: Wang, Zhiwei
; APPLICANT: John Tillinghast
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: No. 6569662el Nucleic Acids and
; FILE REFERENCE: 784CIP2B
; CURRENT APPLICATION NUMBER: US/09/620,312D
; PRIOR FILING DATE: 2000-07-19
; PRIOR FILING DATE: 2000-04-25
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 1105
; SOFTWARE: Pf FL_genes Version 1.0
; SEQ ID NO 1020
; LENGTH: 3324
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (168)..(1625)
US-09-620-312D-1020

Alignment Scores:
Pred. No.: 0.0127 Length: 3324
Score: 129.00 Matches: 100
Percent Similarity: 37.95% Conservative: 81
Best Local Similarity: 20.96% Mismatches: 194
Query Match: 4.16% Indels: 102
DB: 4 Gaps: 22

US-09-890-475-1 (1-609) x US-09-620-312D-1020 (1-3324)
Qy 47 IleGlyGlnSerLys---GlnProGlnPheLeuLysSerIleAsp-----Glu 61
Db 243 CTCCTCAAGAGAGAGTCAAGAGCAACATCTTCTCCACCCAGGGGTCAACTAGAT 302
Qy 62 LeuAlaAlaPheSerValAlaValGluThrPheLysArgGlnPheAspLeuGlnLys 81

Db 303 ATAGCTTCACATCTCTGGATCAAGAAATTTTATTAAGTATAAACTGAATGAAGAA 362
Qy 82 HisIleGluSerIleGluAsnAlaIleAspSerLysLeuGluSerAsnGlyValValLeu 101
Db 363 GAGCTAAATCTCTGGACAAAGAAATTTCTGAAGCCTTCCACGACACAGCCTTT-----416
Qy 102 AlaAlaArgAsnAsnAsnPheHisGlnProMetLeuSerProProArgAsnAsnValSer 121
Db 417 -----GACCGTCACACTTCTCCAGTGTTCAGCCCTGCCAATCCAGAAAGCTCA 464
Qy 122 ValGluThrThrValThr-----ValSerGlnProSer 132
Db 465 ATGGAAGAGCTGCTTGGCCCATCTTGGAGAAAAGTGTCCAGGAAGTGAAGAGCCTCTC 524
Qy 133 GlnGluIleValProGluThrSerAsnLysPro-----GluGlyGlyArgMetCys 149
Db 525 CATAAGCATTGCAAAATGCTCTGAGCCAGCCAGTGCATATCAGGCATTTCCGGAAATGT 584
Qy 150 GluLeu-----MetCysSerLysGlyLeuArgLysTyrIleTyrAlaAsnIleSer 166
Db 585 ACACCTGGAGACCACAGTTCATGCCAGCGCTGGAATAAGATTTTGGTGCTCTGGTTTG 644
Qy 167 AspGlnAlaLysLeuMetGluGluIleProSerAlaLeuLysLeuAlaLysGluProAla 186
Db 645 CTACGCAAAATGCTTTTGAA-----TTGACAAGACTTGTTCAGAACCTTTG 692
Qy 187 LysPheValLeuAspCysIleGlyLysPheTyrLeuGlnGlyArgArgAla-----203
Db 693 AGCGCACTGCTGCAG---TTTGGCGTGACATACCTGGAGGACTATTGCGCAGAGTACATC 749
Qy 204 -----PheThrLysGluSerProMetSerSerAlaArg 214
Db 750 ATTACAAAGTGGTGGGGGCTGTGTTTGTCTTGTAGTCAGAGGAGGAGGAATACCT 809
Qy 215 GlnValSerLeuLeuIleLeuGluSerPheLeuLeuMetProAspArggLyLysGlyLys 234
Db 810 GGAATCACTGCAGAGATAGCAATGACATTTACATCTGCTGCCAGCGACAACCTCGGCAA 869
Qy 235 ValLysIleGluSerTrpLysAspGluAlaGluThrAlaAlaValAlaTrpArgLys 254
Db 870 GTCAGT-----CCCCCAGAGTCTCAACTGTGACCACTTCTCTGCGCAGTCT 914
Qy 255 ArgLeuMetThrGluGlyGlyLeuAlaAlaGluLysMetAspAlaArgGlyLeuLeu 274
Db 915 GAGAGCTTACCTGTG---TCAGTGTGCGGTAGCCAGAGTTGGCACACAGAAAGCTGCCA 971
Qy 275 LeuLeuValAlaCysPheGlyValProSerAsnPheArgSerThrAspLeu---LeuAsp 293
Db 972 GTGTCACTAGGC-----CCTGAGTCTGCGCAGCAGATTGCAATGGATCCCTGAA 1019
Qy 294 LeuIleArgMetSerGlySerAsnGluIleAlaGlyAlaLeuLysArgSerGlnPheLeu 313
Db 1020 GAAGTGAAAGCTTAGACAGCAAC-----GGAGCTGGAGAGAGAGTGAG-----1064
Qy 314 ValProMetValSerGlyIleValGluSerSerIleLysArgGlyMetHisIleGluAla 333
Db 1065 -----AACCAACTCTCTAAATTCGACATTGTGCACCTGGAGAAA 1103
Qy 334 LeuGluMetValTyrThrPheGlyMetGluAspLysPheSerAlaAlaLeuValLeuThr 353
Db 1104 GAAGAGGTG-----CCCAGGGCATGGAGAGGCTGTGTGGCTTCTGTGGTCTTGCCA 1157
Qy 354 SerPheLeuLysMetSerLysGluSerPheGluArgAlaLysArgLysAlaGlnSerPro 373
Db 1158 GCG-----CGGGAGCTGCAAGAGGCACCTTCTCTGAAGCCCGCAGCTCCC 1199
Qy 374 LeuAlaPheLysGluAlaAlaThrLysGlnLeuAlaValLeuSerSerValMetGlnCys 393
Db 1200 TTGCTTCCATATCATCTCCACCTCCCTCGCTGGGACCAAGGAACCTGCACAGAAAGTG 1259
Qy 394 MetGluThrHisLysLeuAspProAlaLys-----GluLeuProGlyTrpGln 409

```

NAME/KEY: CDS
LOCATION: (118)...(3012)
JS-09-620-312D-372

Alignment Scores:
red. No.: 0.0341
core: 130.50
Matches: 137
Conservative: 86
Percent Similarity: 30.80%
Least Local Similarity: 18.92%
Mismatch: 274
Indels: 227
Query Match: 4.20%
JB: 4

JS-09-890-475-1 (1-609) x US-09-620-312D-372 (1-7393)
Y 1 MetSerAsnTyrPro-----ProThrValAlaAlaGlnPro 12
b 859 ATGCCCAACAGTCCCAACTCAAGCCCATTCAGCCCAAGCCCACTGTTATGGAGAACCT 918
Y 13 ThrThrThrAlaAsnProLeu-----GlnArgHisGlnSerGluGlnArg 28
b 919 ---TTCAACAGTCAACCTGCTTGTCTCAGCCCAAGCAGCAAGCAAAAGAAAGCAAA 975
Y 29 ArgArgGluLeuProLysIleValGluThrGluSerThrSerMetAspIleThrIleGly 48
b 976 AAGAAGGAATCTTCAAGGAACCTTGAAGTCTCTGACCCCTCGGAAGGTGTGTGAGCA 1035
Y 49 GlnSerLysGlnProGlnPheLeuLysSerIleAspGluLeuAlaPheSerValAla 68
b 1036 GAGGAAGCAAAAGCCCATTC-----1056
Y 69 ValGluThrPheLysArgGlnPheAspAspLeuGlnLysHisIleGluSerIleGluAsn 88
b 1057 -----AGGAATCTTCAGAAATCGGATGAAATGGAGGGCTCTTAAT 1101
Y 89 AlaIleAspSerLysLeuGluSerAsnGlyValValLeuAlaAlaArgHisAsnAsn 108
b 1102 GGCTCATCAGACCCCAAGCCGCTGGCTAGCATCAAGCTGAAGCCGCAAGATC 1161
Y 109 HisGlnProMetLeuSerProArgAsnValSerValGluThrThrValThrVal 128
b 1162 TACAGTTTCAGCGAATGCCCCCAGCCCTTCATTTGGAGCGAGTAGCCCTTGAAC 1221
Y 129 SerGlnProSerGlnGluIleValPro-----GluThrSerAsnLysProGluGly 145
b 1222 ACTACCCCTACTCAGCCCTGACTCCCTTACATGTGGTGACCCAGATGAGCTGAGCC 1281
Y 146 GlyArgMetCysGluLeuMetCysSerLysGlyLeuArgLysTyrIleTyrAlaAsn 165
b 1282 AGC-----TCAGTCAAAACCAACAGCCCTGCTGATCTCTGACATC 1320
Y 166 SerAspGlnAlaLys-----LeuMetGluGluIleProSer-----AlaLeu 179
b 1321 TCTGATGCTGGGAGGATGGGAGGCGCAAGGTAGACAGTGTCAATCAAGGACGCGAA 1380
Y 180 LysLeuAlaLysGluProAlaLysPheValLeu-----AspCysIleGly 194
b 1381 CAGTTGGTTAAAGAGGGGCTTAAGAAACTTTTTCCTCCCTCAGCCTCAGACCAAGAC 1440
Y 195 LysPheTyrLeuGlnGlyArgAlaPheThrLysGluSer-----208
b 1441 TCACCATATACCAAGGTTTGGAGTTACTATTCTCCAGTTATGTCACAGTCCAGCCCT 1500
Y 209 -----PrometSerSerAlaArgGlnValSerLeuLeuLeuLeuGluSerPhe 224
b 1501 GGGGCTCTGAACCCAGCAGCAGCAGCAAGG-----GTGAGAGCCAG 1542
Y 225 LeuLeuMetProAspArgGlyLysGlyLysValLysIleGluSerTrpIleLysAspGlu 244
b 1543 GCCCTTGAGACAAAGGGATGAGGAACCTCAGACATAGAGGAAGTGAAGAAGCAT 1602
Y 245 AlaGluThrAlaAlaValAlaTrpArgLysArgLeuMetThrGluGlyGlyLeuAla 264
b 1602 -----1602

265 AlaGluLysMetAspAlaArgGlyLeuLeuLeuValAlaCysPheGlyValProSer 284
1603 -----ATCTGTGAAGAAAAGAGGCC 1623
285 AsnPheArgSerThrAspLeuLeuAspLeuIleArgMetSerGlySerAsnGluIleAla 304
1624 GAGCTGAGCAGTTCAGTCCAGAGCCCTCGGTATCCAGCAGCGTCCCAATATGTACATG 1683
305 GlyAlaLeuLysArgSerGlnPheLeuValProMetValSerGlyIleValGluSerSer 324
1684 CAGTCCCTGTACTACCAACCAAGTATGCTATGTACCCCTATGCTACAGGACCAAGAGT 1743
325 IleLysArgGlyMetHisIleGluAlaLeuGluMetValTyrThrPheGlyMetGluAsp 344
1744 TACCAC-----ACCACCTTCTGAGCACTAACCGCTTACCGCAGCAGTACCAAGAA 1797
345 LysPheSerAlaAlaLeuValLeuThrSerPheLeuLysMetSerLysGluSerPheGlu 364
1798 -----CAGCAGAAACGCCAGAGCTTAGAG 1821
365 ArgAlaLysArg-----LysAlaGlnSerProLeuAlaPheLysGluAlaAla 380
1822 CAGCAGCAGCGGGAGTGGACAAGAGGAGCAGATGGGCTTAAGGAGCGGAGGAGCACA 1881
381 ThrLysGln-----LeuAlaValLeuSerSer 389
1882 CTCAGGAAGAGTGAAGCAAAAGCCGTCAATTCACCAACTCTCACCAAGCCCCAGC 1941
390 ValMetGlnCysMetGluThr-----396
1942 CTGACAGACCTGTGAAATCAGGACCTGGCAAGGCCAGGAGCCAGGGGCTGACCCAGCC 2001
397 -----HisLysLeuAspProAlaLysGluLeuProGlyTrpGlnIleLys 411
2002 AATCAGTCTATCTCCAGAGTTAGATGACTCTTCAAACTCCCGGC-----CAGCCCT 2058
412 GluGlnIleValSerLeuGluLysAspThrLeuGlnLeuAspLysGluMetGluGluLys 431
2059 GAAGCCCTTAAAGTGAAGCTGAGTGATGCCAGCACCTAAGCAAGAGGAGCTCTGAGGCC 2118
432 AlaArgSerLeuSerLeuMetGluGluAlaAlaLeuAlaLysArgMetTyrAsnGlnGln 451
2119 AAGCAGGTCTGAGTGTGTGTCAGCAGCAGATGGATCCATCTCTGTGTACGACAG 2178
452 IleLysArgProArgLeuSerProMetGluMetProValThrSerSerSerTyrSer 471
2179 GAGCAGAGCCCCGGATGTGGACATATGTTATCT-----GCCAAGTACTCA 2226
472 ProIle-----TyrArgAspArgSerPheProSerGln 482
2227 GACATCAAGTCAGAGGATGAGCGTGAAGAGGAGCGGAGCCGCAATTTGAAGAGGAA 2286
483 ArgAspAspAspGlnAspGluIle-----490
2287 AGGAGTCGGAGTAGGACTCTGTCCCAAGGAAGATGGGAGGAAGACACAAGTAGTGAC 2346
491 -----SerAlaLeuValSerSerTyrLeuGlyProSerThr 502
2347 TGCAGCTGCCCAAGTCTCAGAGGAGTCTCGCTTTGGAGCAAGAGCCCGGCAAGTGTG 2406
503 SerPhePro-----HisArgSerArgSerProGluTyrMet 515
2407 CATGTGCTGTGCTCTCCCTCCACTTACCAGCAGCAGTCC-----TAC 2448
516 ValProLeuProHisGlyGlyLeuGlyArgSerValTyrAlaTyrGluHisLeuAlaPro 535
2449 ATCCCTTACATGCAGGC-----TATTCTTACAGT-----2478
536 AsnSerTyrSerProGlyHis-----GlyHisArgLeuHisArgGlnTyrSerProSerLeu 554
2479 CAGTCTTACAGACCCCAACCAACCCAGCAGCTACCG-----AGCATGCTGCT 2523

```

QY 226 -----LeuMetProAspArgGly-----LysGlyLysVallys 236
Db 1203 ATCCCATCTGCTCCACTGCTTGACAAAGAGGAGGAGGCGGAAAGCTTGC 1262
QY 237 Ilesluser-----Trp----- 240
Db 1263 ATCTCACCATCTCTCGAGCGCTCTGAAAGTGTGACCGTCCGATGTCTCTCCAC 1322
QY 241 ---IleLysAspGluAlaGluThraLa-----AlaValAlaTrp 252
Db 1323 GTTGTCTGATGAGCAGACACTGCTCTGGGTCTGGTCTACAGCAGTCCGAGAGTGTG 1382
QY 253 ArgLysArgLeuMetThrGluGlyLysLeuAlaAlaGluLysMetAspAlaArgGly 272
Db 1383 AACGAGCCCTCAGGACACAGAGTGTCTATCAGAGCGGAAG-----CGCGT 1433
QY 273 LeuLeuLeuValAlaCysPheGlyValProSerAsnPheArgSerThrAspLeuLeu 292
Db 1434 CTCTGGTACAGGGTGTCTCTCCGA-----GACACTCAG 1469
QY 293 AspleuileArgMetSerGlySer----- 300
Db 1470 GAGCTGTGCGCATCTCGTGTACTCGATACGATGCTCTCAAGTTGTGTGTGAAGCAG 1529
QY 301 -----AsnGluLlealaglyAlaLeu 307
Db 1530 AAGGTCGTCTTTCGCCACCTCAGTCAGTCGGTCTTGTGTCAGCTCAAGGCTTCTCC 1589
QY 308 LysArgSerGlnPheLeuValProMetValSerGlyIleValGluSerSer----- 324
Db 1590 AAGCTCAGCAGACTTGTATTTCGAGGAACAAGTCTGCCCGGAGGCTCTTACACTGCC 1649
QY 325 -----IleLysArgGlyMetHisIleGluAlaLeuGluMetVal 337
Db 1650 CGTCTCTTCCGATGAGAGTAGTCCCGGCCAAGATCATGAGGAGGCTGAGGAGCTC 1709
QY 338 TyrThr-----PheGlyMetGluAspLysPheSerAlaAla 349
Db 1710 TGACACCTCAGACCCCCAGGAATCGCTTGTAGGCTGCCGATCTCTTACTTGTCT 1769
QY 350 LeuValLeuThrSerPheLeuLysMetSerLysGluSerPheGluArgAla----- 366
Db 1770 CTTACCAAGGCGGTGTGTGCGCGGTACTCTTTCGCGATATCGAAAGGAGCCTTGAGCC 1829
QY 366 ----- 366
Db 1830 AAGAGCTGGAAGTCAAGCGCAGGACTGAGATGCTAAGGTAAGTGGCTCAGAAGGAG 1889
QY 367 -----LysArgLysAlaGlnSerProLeuAlaPheLysGluAlaAlaThrLysGlnLeu 384
Db 1890 GGATCAAGCGTGGCGGTCCGTCCGCTGCCACTTCCGCCCTGTCCACCAAGGAGGCC 1949
QY 385 AlaValLeuSerSerValMetGlnCys-----MetGlu 395
Db 1950 GCCCAGAGACCCCTCTGAGAGATCACCATGAGAGCTTCCAGCCCTCCAGGTCTCT 2009
QY 396 ThrHisLysLeuAspProAlaLysGluLeuProGlyTrpGlnIleLysGluGlnIleVal 415
Db 2010 ACCGAGGAGCTCGATGCTGTCTCAAGCGTCTCGCGCAAGAGTCCGATGCCATCTAC 2069
QY 416 SerLeu-----GluLysAspThrLeuGln 423
Db 2070 AAGATCATTTGCCCATCATCGAGAGCTCCGCAAGAACGGCAGCAGGCTGTCTGTCTG 2129
QY 424 LeuAspLysGluMetGluGluLysAlaArgSerLeuSerLeuMetGluGluAlaLeu 443
Db 2130 TACACTCAAGTTCGAGAAGGCTACCTCTTACTAGCCC-CGTCTGAAAGCGCCCTT 2188
QY 444 AlaLysArgMetTyrAsnGlnGlnIleLysArgProArgLeuSerProMetGluMetPro 463
Db 2189 CCCCAAGGAGTTATGC-AGTCCCTGAGGAGACCA---TGTGCGCATCGAGCTGTCT 2244
QY 464 ProValThrSerSerSerTyrSerProIleLysArgAspArgSerPheProSerGln-Ar 483

Db 2245 TCGAACAATCCGAAAGTTCCACGCCGCCAGAGGAGGAGAAC---CCTCCAGGTGG 2301
QY 483 GAspAspAspGlnAspGluIleSerAlaLeuValSerSerTyrLeuGlyProSerThrSe 503
Db 2302 ACACCATCCCGGTGTCTCTGACCGTCTCTCTGTCCTCCATCAGGCGCTCGGTGCT 2361
QY 503 rPheProHisArgSerArgArgSerProGluTyrMetValProLeuProHisGlyLe 523
Db 2362 ACATCCCGGGGTACTCCCGCTTCTCCCGACAC-----TGCCCTTATGCTGGGTG 2412
QY 523 uGlyArgSerValTyrAlaTyrGluHisLeuAlaProAsnSerTyrSerProGlyHisG 543
Db 2413 TTCCCGC-----CATGG 2424
QY 543 yHisArgLeuHisArgGlnTyrSerProSerLeuValHisGlyGlnArgHisProLeuG 563
Db 2425 TCGCGGCTG-----CAACAAGATTGTGTCTG 2451
QY 563 nTyrSerProPro-----IleHisGlyGlnGlnLeuPro----- 575
Db 2452 CTCTCTCTCCCGCGCCAGCAACCATCATCTCCGAGATTGTCCAGTCCGCTCACAAGG 2511
QY 576 -TyrGly-----IleGlnArgValTyrArgHisSerProSer----- 587
Db 2512 TTGGGCGCGAGTCCATCTGTCTGCGCGGTGCCAGCGCTAGCTGCCATGCGCTAAG 2571
QY 588 ----GluGluArgTyrLeuGlyLysLeuSerAsnGlnArgSerProArgSerAsnSerSe 606
Db 2572 GCACCGAGAGCATCAACCAAGTCGCAAGATTCTCGGCC-CCGTAACCAAGTTCGTCT 2630
QY 606 uAspProLys 609
Db 2631 GCTGCCAAGA 2640

RESULT 8

US-09-620-312D-372
; Sequence 372, Application US/09620312D
; Patent No. 6569662
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyan
; APPLICANT: Chen, Rui-hong
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wehrman, Tom
; APPLICANT: Xue, Aidong J.
; APPLICANT: Yang, Yonghong
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Zhou, Ping
; APPLICANT: Ma, Yungling
; APPLICANT: Wang, Dunrui
; APPLICANT: Wang, Zhiwei
; APPLICANT: John Tillinghaast
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: No. 6569662el Nucleic Acids and
; TITLE OF INVENTION: Polypeptides
; FILE REFERENCE: 784CIP2B
; CURRENT APPLICATION NUMBER: US/09/620,312D
; CURRENT FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/488,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 1105
; SOFTWARE: pt_FL_genes Version 1.0
; SEQ ID NO 372
; LENGTH: 7393
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:


```

APPLICATION NUMBER: 08/699,052
FILING DATE: 16-AUG-1996
ATTORNEY/AGENT INFORMATION:
NAME: Glaister, Debra J
REGISTRATION NUMBER: 33,888
REFERENCE/DOCKET NUMBER: GC361-2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-846-7620
TELEFAX: 650-845-6504
INFORMATION FOR SEQ ID NO: 29:
SEQUENCE CHARACTERISTICS:
LENGTH: 17612 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-911-853-29

Alignment Scores:
Pred. No.: 0.0814 Length: 17612
Score: 133.00 Matches: 131
Percent Similarity: 35.95% Conservative: 98
Best Local Similarity: 20.57% Mismatches: 224
Query Match: 4.28% Indels: 184
DB: 3 Gaps: 31

US-09-890-475-1 (1-609) x US-08-911-853-29 (1-17612)
QY 14 ThrThrAlaAsnProLeuLeuGlnArgHisGlnSerGluGlnArgArgGluLeuPro 33
DB 9754 ACCTCGATCTGCGCGTACCAATCGTAGTCCGCGCAGCAACGGCGCAACAGCGCGGTG 9813
QY 34 LysIleValGluThrGluSerThrSerMetAspIleThrIleGlyGlnSerLysGlnPro 53
DB 9814 AAGATCGCGCAGATGAGTGCACCTCGCGCTCATCGCC----- 9852
QY 54 GlnPheLeuLysSerIleAspGluLeuAlaAlaPheSerValAlaValGluThrPheLys 73
DB 9853 -----TCGGCCTTCGGCGACACGGGGGTGTTCTTCAAG 9885
QY 74 ArgGlnPheAspLeuGlnLysHisIleGluSerIleGluAsnAlaIleAspSerLys 93
DB 9886 CAC-----CAGCGTCACGAGCAATAAGAGGTGTCATCCCTTGCTGAACAG 9930
QY 94 LeuGluSerAsnGlyValValLeuAlaAlaArgAsnAsn-----PheHisGln 110
DB 9931 CCGCGCGCGTGGCGGGCTTTTGTGCACGCTTACGTCATCACATTCACCTTCGGCGCAG 9990
QY 111 ProMetLeuSerProProArgAsnAsnValSerValGluThrValThrValSerGln 130
DB 9991 GCTGTGTCGCGCT-----GCAAAATCGGCACCTGCAGTTTTTTGGCG 10032
QY 131 ProSerGlnGluLeuValProGluThrSerAsnLysProGluGlyArgMetCysGlu 150
DB 10033 AATTCCTTAATTGGCGCTCGGCCATCCCATTAATAAACAACAGACACAGCAAGATG 10092
QY 151 -LeuMetCysSerLysGlyLeuArgLysTyrIle-----TyrAlaAsnIleSerAspG 168
DB 10093 GATCTTCTGTTGGGGAACGATCCGCCATGTCACCGATACCCACGCGCGCTGACGG 10152
QY 168 nAlaLysLeuMetGluGluLeuProSerAlaLeuLysLeuAlaLysGluProAlaLys-P 188
DB 10153 CTCCCGCAAGCCCGCTTCGCGCGCGCTGCGCTTCGCTTCGCAACGCCACCGCGTGC 10212
QY 188 heValLeuAspCysIleGlyLysPheTyrLeuGlnGlyArgAlaPheThrLysGlu 208
DB 10213 TGCTGCGCGAGCCCTTCGGCGCAGTCCAGCTGCAGGTGCGCGCGT----- 10259
QY 208 erProMetSerSerAlaArgGlnValSerLeuLeuIleLeuGluSerPheLeuLeuMetP 228
DB 10260 -----GCCAGCCTGCCCGCGTGCAGGAG-----G 10284
QY 228 roAspArgGlyLysGlyValLysIleGluSerTrpIleLysAspGluAla-----G 246
DB 10284 -----G 246

10285 CCACAGCGCTTCGCGCGCGCGTGTGCGCTGCACCTGCCTGGAGCCCGAGCCCTTCGAGC 10344
QY 246 LuThrAlaAlaValAlaTrpArgLys-----ArgLeuMetThrGluGlyG 261
DB 10345 AGGAGCTGCGCCCTGCGCTTACGAGCGCGACTCTCCGAGGTGCGGCAGATGCGCGAGGC- 10403
QY 261 LysLeuAlaAlaGluLysMetAspAlaArgGlyLeuLeuLeuValAlaCysPheG 281
DB 10404 -----ATGGGTCCGAACTTGCACCTAGCAGCTTGGCGCACTCACTCCC----- 10448
QY 281 LysValProSerAsnPheArgSerThrAspLeuLeuAspLeuIleArgMetSerGlySerA 301
DB 10449 -----GAATCCGCGCGACCTGCTGGAG----- 10469
QY 301 snGluIleAlaGlyAlaLeuLysArgSerGlnPheLeuValProMetValSerGlyIleV 321
DB 10470 -----CAGGAAGATGACGCGCGCATCATCGCGCTGATCAACGCCATCC 10512
QY 321 alGluSerSerIleLysArgGly-----MetHisIleGluAlaLeuGlu-----M 336
DB 10513 TCAGCGAGCGCATCAAGCGCGCGCTCCGACATCCACCTCGAAACCTTCGAGAAACGCC 10572
QY 336 etValTyrThrPheGlyMetGluAspLysPheSerAlaAlaLeuValLeuThrSerPheL 356
DB 10573 TGGTGTGCGCTTTCGCGTCGACGCG-----ATCCTCGCGAAGTGA 10614
QY 356 euLysMetSerLysGluSerPheGluArgAlaLysArgLysAlaGlnSerProLeuAlaP 376
DB 10615 TCGAACCGCGCGCGAG-----CTGCGC- 10637
QY 376 heLysGluAlaAlaThrLysGlnLeuAlaValLeuSerSerValMetGlnCysMetGluT 396
DB 10638 -----GGCTGTGCTGTTCGCGGTCAAGGTCAATGCGG- 10670
QY 396 hrHisLysLeuAspProLa-----LysGluLeuPro----- 406
DB 10671 -----CGCTGTGACATCGCGAGAGCGGTACGCGAGCGCGCGTATTTCGCTCAAGG 10725
QY 407 -----GlyTrpGlnIleLysGluGlnIleValSerLeu-----GluLysA 420
DB 10726 TCGCGGTGCGAGGTGATATCCCGTCTCCACCTGCGCTCGGCCAACCGCGAGCGGG 10785
QY 420 spThrLeuGlnLeuAspLysGluMetGluLysAlaAlaArgSerLeuSerLeu----- 437
DB 10786 TGTGTGCTGCTG-----CTCGACAAGAGCGCGCGCGCTGTGCTCACGCATC 10836
QY 438 -----MetGluGluAlaAlaLeuAlaLysArgMetTyrAsnGlnGlnIleLysArgPro- 455
DB 10837 TGGGCATGAGCGAGCGC-----GACCGCGCGCTGCTCGAGCAACCTGCGCAAGCGCG 10890
QY 456 -ArgLeuSerProMetGluMetProValThrSerSerSerTyrSerProIleTyrAr 475
DB 10891 ACGGCATCATCTAGTACCGCGCGCGCGCTCGGCGAAGACCAACCACTGTACGCCG 10950
QY 475 gasPArgSerPheProSerGlnArgAspAspAspGlnAspGluLeuSerAlaLeuValse 495
DB 10951 GCCT---GGTCAACCTCAACGA----- 10969
QY 495 rSerTyrLeuGlyProSerThrSerPheProHisArgSerArgArgSerProGluTyrMe 515
DB 10970 -----CGCTCGCGCATATCTTCACGGTGAAGA-----CCCGATCGAGT 11010
QY 515 tValPro-----LeuProHisGlyGlyLeuGlyArgSerVa 527
DB 11011 ACTACCTGAAGGATCGGCCAGACCCAGGTCAACCGCGGTGAGATGACCTTCGCCCC 11070
QY 527 lTyrAlaTyrGluHisLeuAlaPro-----AsnSerTyrSerProG 541
DB 11071 GCGCGCTGCGCGCATCTTCGCCAGGACCGGACGCTGATGTCGGCGAGATCCGCG 11130
QY 541 yHisGlyHisArgLeuHisArgGlnTyrSer-----ProSerLeuVal-- 555
DB 11131 ACCAGGAGACCGCGACATCGCGGTGCGAGCGCTCGCTCACCGGCGCACCTGTGCTCTCCA 11190

```



```
QY 318 SerGlyIleValGluSerSerIleLysArgGlyMetHisIleGluAlaLeuGluMetVal 337
Db 2209 -----GAACGGAACTAAGACGG-----AAACATGTC----- 2235
QY 338 TyrThrPheGlyMetGluAspIlePheSerAlaLeuValLeuThrSerPheLeuLys 357
Db 2236 -----ATGAAGTTGCACAGCGCTAGAGTTGAAGTCTAAAGACTCA 2283
QY 358 MetSerLysGluSerPheGluArgAlaLysArgLysAlaGlnSerProLeuAla 375
Db 2284 ATAAAAAGCAGTTTCAGGATACCTGCAAAATTCAAACACAGACAGTACAAAGCATTAAAG 2343
QY 376 -----PheLys 377
Db 2344 AATCACTACTGAGACTACACAAAGAGTGAACAAAGCTGTCTGAAAAGACTCAAG 2403
QY 378 GluAlaAlaThrLysGlnLeuAlaValLeuSer-----SerValMetGln 392
Db 2404 GAGGAACAGACTCGGAAGTTAGCCATCTTGGCTGAGCAGTATGATCATAGCATTAATGAA 2463
QY 393 CysMetGluThrHis-----LysLeuAspProAlaLysGlu----- 404
Db 2464 ATGCTCTCCACACAGACTCTGCGTTGGATGAAGCACAGAGCAGAGTGCAGGTTTGG 2523
QY 405 -----LeuProGlyTyrGlnIleLys----- 411
Db 2524 AAGATGACGCTACAGCAGGAAGCTGGAGCTGTGGAATGCATACAGACAAATCAAGATG 2583
QY 412 -----GluGlnIleValSerLeu----- 417
Db 2584 CAGGCTGAGGCCCAACATGATGAGAGCTTCGAGAGCTGGAACAAAGGCTCTCCCTTCGG 2643
QY 418 -----GluLysAspThrLeuGlnLeuAspLysGluMetGlu 429
Db 2644 AGAGCACTCTTAGACAGAGATTGAAGAGAGATGTTGGCTTTCAGAAATCAAGCACA 2703
QY 430 GluLysAlaArgSerLeuSerLeuMetGluGluAlaLeuAlaLysArgMetTyrAsn 449
Db 2704 GAACGATACGT-----AGCCTGCTCGAGCGCCAGAGAAATTAAGACTTTTGAC 2757
QY 450 GlnGlnIleLysArgProArgLeuSerProMetGluMetProProValThrSerSerSer 469
Db 2758 TCTGAAGCATGATGATTAGTTTGTATACATGCTCTT-----TCTAAT 2802
QY 470 TyrSerProIleTyrArgAspArgSerPheProSerGlnArgAspAspGlnAspGlu 489
Db 2803 CTCTCCCTCGAGCATTCAGCCACAGTACCCAGGA----- 2838
QY 490 IleSerAlaLeuValSerSerTyrLeuGlyProSerThrSerPheProHisArgSerArg 509
Db 2839 -----GCTTCTAGTGTCTCACAAT----- 2859
QY 510 ArgSerProGluTyrMetValProLeuProHisGlyGlyLeuGlyArgSerValTyrAla 529
Db 2860 -----CCTACTGGGGGTTTCAGGA----- 2877
QY 530 TyrGluHisLeuAlaProAsnSerTyrSerProGlyHisGlyHisArgLeuHisArgGln 549
Db 2878 -----CCTCACTGGGGTCACTCCCATG----- 2898
QY 550 TyrSerProSerLeuValHisGlyGlnArgHisProLeuGlnTyrSerPro----- 566
Db 2899 -----GCTGSCACACCAAGCTTGGGTCATCCGATGCAAGGCGGACCCCAACCATGG 2952
QY 567 -----ProIleHisGlyGlnGlnGlnLeuProTyrGlyIleGlnArgValTyrArgHis 584
Db 2953 GGTCAACCCCTCAGGGCCCAATGCAAGGGGTACCTCGAGGTAGCATGATAGGATCGCAAT 3012
QY 585 SerProSerGluGluArgTyrLeu-----GlyLeuSerAsnGlnArgSerProArg 601
Db 3013 AGCCCCCAGGCTCTGAGGCGGACAGCTTCTGGGGGACGAGCGGACCAACAGGCGCATGAGCAG 3072
QY 602 SerAsnSer 604
```

```
Db 3073 AGCAGGACT 3081
RESULT 3
US-09-491-356C-7
; Sequence 7, Application US/09491356C
; Patent No. 6566061
; GENERAL INFORMATION:
; APPLICANT: Philibert, Robert A.
; APPLICANT: Ginns, Edward I.
; TITLE OF INVENTION: IDENTIFICATION OF POLYMORPHISMS IN THE PCTG4 REGION OF XQ13
; FILE REFERENCE: 9465.6US11
; CURRENT APPLICATION NUMBER: US/09/491.356C
; CURRENT FILING DATE: 2000-01-26
; PRIOR APPLICATION NUMBER: PCT/US99/09365
; PRIOR FILING DATE: 1999-04-29
; PRIOR APPLICATION NUMBER: 60/083,465
; PRIOR FILING DATE: 1998-04-29
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7
; LENGTH: 6558
; TYPE: DNA
; ORGANISM: Mus musculus
US-09-491-356C-7
Alignment Scores:
Pred. No.: 0.0055 Length: 6558
Score: 137.50 Matches: 137
Percent Similarity: 33.45% Conservative: 118
Best Local Similarity: 17.98% Mismatches: 226
Query Match: 4.43% Indels: 281
DB: 4 Gaps: 36
US-09-890-475-1 (1-609) x US-09-491-356C-7 (1-6558)
QY 17 AnProLeuLeuGlnArgHisGlnSerGluGlnArgArgGluLeuProLysIleVal 36
Db 3928 GACCCAGTGTGAGTAGTCCAGGCCCGCCGCGC-----CTCATGCAACTATC 3975
QY 37 GluThrGluSerThrSerMetAspIleThrIleGlyGlnSerLysGlnProGln----- 54
Db 3976 TGCTACCCACATCGACTGCTGCAATCAGGATGAGAAACCCCGAGCGCGCGCAT 4035
QY 55 -----PheLeuLysSerIleAspGluLeuAlaPheSerValAlaValGluThrPhe 72
Db 4036 AAACGTAATCTCAAGAAATTAGCAATGG-----ACCATG 4071
QY 73 LysArgGlnPheAspAspLeuGlnLysHisIleGluSerIleGluAsnAlaIleAspSer 92
Db 4072 CGCCAGTCTCTTTGGAGCTACAGTTGATGATCAAGCAGACCCCAACACT----- 4122
QY 93 LysLeuGluSerAsnGlyValValLeuAlaAlaPheSerValAlaValGluThrPhe 112
Db 4123 -----GAGATGAATCTCTCTTG----- 4140
QY 113 LeuSerProArgAsnAsnValSerValGluThrValThrValThrValSerGlnProSer 132
Db 4141 -----GAGACATGTC-----AAGGCCACAAATCGAGTTTCCACAGTCA 4182
QY 133 GlnGluIleValProGluThr-----SerAsnLysProGluGlyGlyArgMet 148
Db 4183 GCAGACAGCGGTTCATCTTCTGGAAGTACAGCAACATGCCCCGAGCAGCAAGACC 4242
QY 149 CysGluLeuMetCysSerLysGlyLeuArgLysTyrIleTyrAlaAsnIleSerAspGln 168
Db 4243 AAACCTGTGCTCAGTCT-----CTAGAGCGA-----TCTGGTGTATGTTGGTG 4287
QY 169 AlaLysLeuMetGlnGluIleProSerAlaLeuLysGluProAlaLysPhe 188
Db 4288 GCTCTCTCATTCGCAAACTGCCCACTTCAGTCCAG-----GGCCAT 4329
```

2878 -----CCTCAGTGGGGTCAATCCCATG----- 2898
 550 TyrSerProSerLeuValHisGlyGlnArgHisProLeuGlnTyrSerPro----- 566
 2899 -----GGTGGCACACACCAAGCTTGGGTGATCCGATGCAAGCGGACCCCAACCATGG 2952
 567 -----ProTleHisGlyGlnGlnLeuProTyrGlyGlnArgValTyrArgHis 584
 2953 GGTCAACCCCTCAGGCGCAATGCAAGGGTACCTCGAGGTAGCATGATAGAGTCCGCAAT 3012
 585 SerProSerGluGluArgTyrLeu-----GlyLeuSerAsnGlnArgSerProArg 601
 3013 AGCCCCCAGGCTCTGAGCGGACAGCTTCTGGGGGACGCGACGCAAGCGGCATGACGACA 3072
 602 SerAsnSer 604
 3073 AGCAGCAGT 3081

RESULT 2

JS-09-723-458-1
 : Sequence 1, Application US/09723458
 : Patent No. 6586242
 : GENERAL INFORMATION:
 : APPLICANT: Cobb, Melanie
 : Hutchinson, Michele
 : Chen, Zhu
 : Berman, Kevin
 : TITLE OF INVENTION: TAO PROTEIN KINASES AND METHODS OF USE
 : THEREFOR

NUMBER OF SEQUENCES: 26
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: SEED and BERRY LLP
 STREET: 6300 Columbia Center, 701 Fifth Avenue
 CITY: Seattle
 STATE: Washington
 COUNTRY: USA
 ZIP: 98104

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent in Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/723,458
 FILING DATE: 27-No. 6586242-2000
 CLASSIFICATION: <Unknown>
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US/09/060,410
 FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:
 NAME: Maki, David J.
 REGISTRATION NUMBER: 31,392
 REFERENCE/DOCKET NUMBER: 860098.421
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (206) 622-4900
 TELEFAX: (206) 692-6031
 INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:
 LENGTH: 3312 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 FEATURE:

NAME/KEY: CDS

LOCATION: 121..3123

SEQUENCE DESCRIPTION: SEQ ID NO: 1:
JS-09-723-458-1

Alignment Scores:

Pred. No.: 7,64e-05 Length: 3312
 Score: 151.00 Matches: 150
 Percent Similarity: 34.75% Conservative: 95
 Best Local Similarity: 21.28% Mismatches: 237

Query Match: 4.86% Indels: 223
 DB: 4 Gaps: 32
 US-09-890-475-1 (1-609) x US-09-723-458-1 (1-3312)
 QY 3 AsnTyr-----ProProThrValAlaAlaGlnProThrThrThrAlaAsn 17
 DB 1333 AATTCCAGAGAAGAGATCTTAGAACAGAGATCMGCTCCACAGTCTCCA----- 1386
 QY 18 ProLeuLeuGlnArgHisGlnSerGluGlnArgArgGluLeuProLysIleValGlu 37
 DB 1387 CCTCAAGTGCTCGTCACAAATCACATTATCGTAATAGAGAACTTTGCAACTATAGA 1446
 QY 38 ThrGluSerThrSerMetAspIleThrIleGlyGlnSerLysGlnProGlnPheLeuLys 57
 DB 1447 ACAGCATCACTG-----GTTACAAGACAGATGCAAGAACATGAG 1485
 QY 58 SerIleAspGluLeuAlaAlaPheSerValAlaValGluThrPheLysArgGlnPheAsp 77
 DB 1486 CAGGACTCTGAACCTTAGAGAACAGATGTCTGTTATTAAGCGGATGAGCGACAGCAT--- 1542
 QY 78 AspLeuGlnLysIleGluSerIleGluAsnAlaIleAspSerLysLeuGluSerAsn 97
 DB 1543 -----CAGAACGACGTGATGACTCTGGAATAATAAACTGAAGCGCAAAATGACGAACAT 1596
 QY 98 GlyVal-----ValLeuAlaAlaArgAsnAsnPheHisGlnProMetLeu 113
 DB 1597 CGGCTCAGATTAGACAAGATCTTGAACCTCAGCGCAACAATTCGCTGCAGAAATG--- 1653
 QY 114 SerProProArgAsnAsnValSerValGluThrThrValThrValSerGlnProSerGln 133
 DB 1654 -----GAGAACTTATTAGAAGAACACCAAGCTTCTATG 1686
 QY 134 GluIleValProGluThrSerAsnLysProGluGlyGlyArgMetCysGluLeuMetCys 153
 DB 1687 GAAAAGAGGCTAAAGTATGCGCCACAGAGGAG----- 1719
 QY 154 SerLysGlyLeuArgLysTyrIleTyrAlaAsnIleSerAspGlnAlaLysLeuMetGlu 173
 DB 1720 -----AAAAAATCCAAACACAT-----CAGGCTCACAGAGAGAA 1758
 QY 174 GluIleProSerAlaLeuLys-LeuAlaLysGluProAlaLysPheValLeuAspCysIle 193
 DB 1759 GAACGTGAATAGCTTTTGGAGTCTCAAAAAGAGAAATATAAACTTCGAAAAGAGAGAGCTT 1818
 QY 193 eGlyLysPheTyrLeuGlnGlyArgAlaPheThrLysGluSer-----ProMetSe 211
 DB 1819 AAGGAGGAGCTGAATGAAACCCAGACACACCTTAAAAAAGAAAGCAGGATGGCTTTCA 1878
 QY 211 rSerAlaArgGlnValSerLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 231
 DB 1879 AAGCAGAAGAGAGATATTCAACATTTTCAGGCGAGAGAGAGAGCTAATCTTCTTCAGCT 1938
 QY 231 YLysGlyLysVal-----LysIleGluSerTrpIleLysAspGluAlaGluTh 247
 DB 1939 CAAGGAGTATCTAGAGCTAGATGTCGCTGCTTCA-----AAAGAAGATGTTACTT 1992
 QY 247 rAlaAlaValAlaTrpArgLysArgLeuMetThrGlu-GlyGlyLeuAlaAlaGluL 267
 DB 1993 GGTGGCATAACTTGAACAGGACCTTG-----TCAGGAGGAGGTT-----AAACAAA 2040
 QY 267 yMetAspAlaArgGlyLeuLeuLeu-----LeuValAlaLac 279
 DB 2041 AGCAGACTCAGAGAGGACTTAGAACATGCAATGTTACTCGGACAGCATGAATCCATGCA 2100
 QY 279 yspheGlyVal---ProSerAsnPheArgSerThrAspLeuLeu-AspLeuIleArgMet 297
 DB 2101 GAACGTGAGTTTCGCCACCTCAACACTATTCAAGAGATCGCTGTGAGTTGATCAGACTG 2160
 QY 298 SerGlySerAsnGluIleAlaGlyAlaLeuLysArgSerGlnPheLeuValProMetVal 317
 DB 2161 CAACATCAAACTGAGCTTACTTAACAGCTGGGAATACATAAGAGAGAGG----- 2208

SEQUENCE CHARACTERISTICS:

; LENGTH: 3312 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: 121..3123
 ; US-09-060-410-1

Alignment Scores:

Pred. No.: 7,64e-05 Length: 3312
 Score: 151.00 Matches: 150
 Percent Similarity: 34.75% Conservative: 95
 Best Local Similarity: 21.28% Mismatches: 237
 Query Match: 4.86% Indels: 223
 DB: 3 Gaps: 32

US-09-890-475-1 (1-609) x US-09-060-410-1 (1-3312)

QY 3 AsnTyr-----ProThrValAlaAlaGlnProThrThrAlaAsn 17
 DB 1333 AATTACCAAGAGAGAGATCTAGACAAAGAGCATCAGCTCCACAGTCTCCA-----1386
 QY 18 ProLeuLeuGlnArgHisGlnSerGluGlnArgArgGluLeuProLysileValGlu 37
 DB 1387 CCTCAAGTGTCTCTCACAATACATTCATCTGTAATAGAGACACTTTGCCAATATACGA 1446
 QY 38 ThrGluSerThrSerMetAspIleThrIleGlyGlnSerLysGlnProGlnPheLeuLys 57
 DB 1447 ACAGCATCTG-----GTTACAGACAGATGAGCGGCAATGAG 1485
 QY 58 SerIleAspGluLeuAlaPheSerValAlaValGluThrPheLysArgGlnPheAsp 77
 DB 1486 CAGGACTCTGAATCTAGAACACAGATGCTGTTATAGCGGATGAGCGGACAGCAT---1542
 QY 78 AspLeuGlnLysHisIleGluSerIleGluAsnAlaIleAspSerLysLeuGluSerAsn 97
 DB 1543 -----CAGAAAGCAGCTGATGATCTCTGGAATATAAATCTGAAGCGAGAAATGAGCAACAT 1596
 QY 98 GlyVal-----ValleuAlaAlaArgAsnAsnAsnPheHisGlnProMetLeu 113
 DB 1597 CGGCTCAGATTAGAACAAAGATCTTGAATCTAGCCGACAAATTTCTGTCGAGAAATG---1653
 QY 114 SerProArgAsnValSerValGluThrThrValThrValSerGlnProSerGln 133
 DB 1654 -----GAGAAATTTATTAAAGAAACACCAAGCTTCTATG 1686
 QY 134 GluIleValProGluThrSerAsnLysProGluGlyGlyArgMetCysGluLeuMetCys 153
 DB 1687 GAAAAAGAGCTAAAGTGTAGCCCAACGAGGAG-----1719
 QY 154 SerLysGlyLeuArgLysTyrIleTyrAlaAsnIleSerAspGlnAlaLysLeuMetGlu 173
 DB 1720 --AAAAAATTCACACACATTT-----CAGGCTCAACAGAGAGAA 1758
 QY 174 GluIleProSerAlaLeuLys-LeuAlaLysGluProAlaLysPheValLeuAspCysI1 193
 DB 1759 GAACTGAATAGCTTTTGGAGTCTCAAAAAGAGAGATATAAACTTCGAAAAAGACAGCTT 1818
 QY 193 eGlyLysPheTyrLeuGlnGlyArgAlaPheThrLysGluSer-----ProMetSe 211
 DB 1819 AAGGAGGAGCTGAATGAACACAGACACACCTAAAAAGAAAGACGAGGATGGCTTCA 1878
 QY 211 rSerAlaArgGlnValSerLeuLeuIleLeuGluSerPheLeuLeuMetProAspArgG1 231
 DB 1879 AAGCAGAGAGAGAGATATCAACATTTTCAGGCGAGAGAGAGAGCTAATCTTCTCGACGT 1938
 QY 231 YLysGlyLysVal-----LysIleGluSerTrpIleLysAspGluAlaGluTh 247
 DB 1939 CAAGGAGCATCTAGAGCTAGATGCTGCTGCTCA-----AAAGAGAGATGTACTT 1992
 QY 247 rAlaAlaValAlaTrpArgLysArgLeuMetThrGlu-GlyGlyLeuAlaAlaGluL 267

DB 1993 GGTGGCATAAAGTGGAAACAGGACCTTG-----TCAGGAGGAGTT-----AAACAA 2040
 QY 267 yMetAspAlaArgGlyLeuLeuLeu-----LeuValAlaC 279
 DB 2041 AGGAGACTCAGAAGGACTTGAACATGCAATGTTACTGCGACAGCATGATCATCGTCAA 2100
 QY 279 yPheGlyVal----ProSerAsnPheArgSerThrAspLeuLeu-AspLeuIleArgMet 297
 DB 2101 GAACTGGAGTTTGGCCACCTCAACACTATTCAAGAGATGCGCTGTGAGTTGATCAGACTG 2160
 QY 298 SerGlySerAsnGluIleAlaGlyAlaLeuLysArgSerGlnPheLeuValProMetVal 317
 DB 2161 CAACATCAAACTGAGCTTACTTAACAGCTGGAATPACAAATAGAGAGG-----2208
 QY 318 SerGlyIleValGluSerSerIleLysArgGlyMetHisIleGluAlaLeuGluMetVal 337
 DB 2209 -----GAACGGGAACTAAGACGG---AAACATGTC-----2235
 QY 338 TyrThrPheGlyMetGluAspLysPheSerAlaAlaLeuValLeuThrSerPheLeuLys 357
 DB 2236 -----ATGGAAGTTTCGACAGCAGCCTAAGAGTTTGAAGTCTAAAGACTCCAA 2283
 QY 358 MetSerLysGluSerPheGluArgAlaLysArgLysAlaGlnSerProLeuAla-----375
 DB 2284 ATAAAAAGCAGTTTTCAGGATACCTGCAAAATTCAAACACAGACAGTACAAAGCATTAAGG 2343
 QY 376 -----PheLys 377
 DB 2344 ATACCTACTGGAGACTACCAAGAGTGAGCACAAGCTGTTCTGAAAAGACTCAAG 2403
 QY 378 GluAlaAlaThrLysGlnLeuAlaValLeuSer-----SerValMetGln 392
 DB 2404 GAGGAACAGACTCGGAAGTTAGCCATCTTGCTGAGCAGTATGATCATGATTAATGAA 2463
 QY 393 CysMetGluThrHis-----LysLeuAspProAlaLysGlu-----404
 DB 2464 ATGCTCTCCACACAAGCTCTGCTGTTGATGAACACAGAGAGCAGAAATCCAGGTTTG 2523
 QY 405 -----LeuProGlyTrpGlnIleLys-----411
 DB 2524 AAGATGCAGCTACACAGCAAGTGGAGCTGTTGAATGATATATCAGAGCAAAATCAAGATG 2583
 QY 412 -----GluGlnIleValSerLeu---417
 DB 2584 CAGGCTGAGGCCCAACATGATCGAGAGCTTCGAGAGCTGGAAACAAAGGGTCTCCCTTCGG 2643
 QY 418 -----GluLysAspThrLeuGlnLeuAspLysGluMetGlu 429
 DB 2644 AGAGCACTCTTAGACAGAAAGATTGAAGAGAGATGTTGGCTTTCAGAGATGAACGCACA 2703
 QY 430 GluLysAlaArgSerLeuSerLeuMetGluGluAlaAlaLeuAlaLysArgMetTyrAsn 449
 DB 2704 GAACGAATACGT-----AGCCTGCTCGAGCGCCAGAGAAATGAAGCTTTTGAC 2757
 QY 450 GlnGlnIleLysArgProArgLeuSerProMetGluMetProProValThrSerSerSer 469
 DB 2758 TCTGAAGCATGAGATTAGTTTATGTAACATGTCCTT-----TCTAAT 2802
 QY 470 TyrSerProIleTyrArgAspArgSerPheProSerGlnArgAspAspGlnAspGlu 499
 DB 2803 CTCTCCCTCAGGCATTTCAGCACAGCTTACCAGGA-----2838
 QY 490 IleSerAlaLeuValSerSerTyrLeuGlyProSerThrSerPheProHisArgSerArg 509
 DB 2839 -----GCTTCTAGCTGTCTCAAT-----2859
 QY 510 ArgSerProGluTyrMetValProLeuProHisGlyGlyLeuGlyArgSerValTyrAla 529
 DB 2860 -----CCTACTGGGGGTTCCAGGA-----2877
 QY 530 TyrGluHisLeuAlaProAsnSerTyrSerProGlyHisGlyHisArgLeuHisArgGln 549

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

DM protein - nucleic search, using frame_plus_p2n model
Run on: August 25, 2003, 10:24:14 ; Search time 104 Seconds
(without alignments)
2584.638 Million cell updates/sec

Title: US-09-890-475-1
Perfect score: 3104
Sequence: 1 MSNPPVAAQPTTANPLL.....RYLGSLNQRSPRNSLDPK 509

Scoring table: BLOSUM62

Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Dgapop 6.0, Delcxt 7.0

Searched: 559978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool/US09890475/runat 14082003 090351 19610/app query.fasta_1.775
-DB=Issued Patents NA -QWMI=fastcap -SUFFIX=rni -MINMATCH=0.1 -LOOPTCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=BLOSUM62 -TRANS=human0.cdi
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09890475 @CGN 1.1 56 @runat 14082003 090351 19610 -NCPU=6 -ICPU=3
-NO WMAP -LARGESQUERY -NEG SCORES=0 -WAIT -DSFBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents NA:*

1: /cgn2_6/ptodata/2/ina/5A COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PTUS COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	151	4.9	3312	3	US-09-060-410-1
2	151	4.9	3312	4	US-09-723-458-1
3	137.5	4.4	6558	4	US-08-431-356C-7
4	133	4.3	17612	3	US-08-911-853-29
5	133	4.3	17612	3	US-09-479-409-29
6	133	4.3	17612	4	US-09-479-453-29
7	131	4.2	9775	3	US-08-977-171-1
8	130.5	4.2	7393	4	US-09-620-312D-372
9	129	4.2	3324	4	US-09-620-312D-1020
10	127.5	4.1	3605	3	US-09-098-901-1
11	127	4.1	6306	1	US-08-466-390-3
12	127	4.1	6306	1	US-08-470-950-3

13	127	4.1	6306	1	US-08-467-781-3
14	127	4.1	6306	1	US-08-195-487-3
15	127	4.1	6306	1	US-08-483-924-3
16	127	4.1	6306	5	PCT-US93-06160-3
17	126.5	4.1	9934	3	US-08-977-171-2
18	125.5	3.9	7100	4	US-09-308-375-1
19	120	3.9	11220	3	US-09-105-537-32
20	120	3.9	36778	3	US-09-105-537-5
21	119.5	3.8	5053	2	US-08-685-576-2
22	118	3.8	38506	3	US-09-320-878-19
23	118	3.8	38506	4	US-09-141-908-1
24	118	3.8	38506	4	US-09-657-440-19
25	117.5	3.8	2487	1	US-08-261-304-1
26	116.5	3.8	4403765	3	US-09-103-840A-2
27	116.5	3.8	4411529	3	US-09-103-840A-1
28	116	3.7	1512	3	US-08-911-853-6
29	116	3.7	1512	3	US-09-479-409-6
30	116	3.7	1512	4	US-09-479-453-6
31	116	3.7	2680	2	US-08-533-306A-5
32	116	3.7	2680	2	US-08-742-923A-5
33	116	3.7	2887	2	US-08-533-306A-3
34	116	3.7	2887	2	US-08-742-923A-3
35	116	3.7	7453	4	US-09-620-312D-248
36	116	3.7	7501	4	US-09-620-312D-249
37	116	3.7	10136	1	US-08-353-700-2
38	116	3.7	10136	5	PCT-US95-16216-2
39	115.5	3.7	2859	4	US-09-252-991A-944
40	115	3.7	3378	4	US-09-513-783A-151
41	115	3.7	4833	4	US-09-513-783A-21
42	115	3.7	8789	1	US-08-328-254-5
43	114.5	3.7	1836	4	US-09-252-991A-15831
44	114.5	3.7	2055	4	US-09-252-991A-15771
45	114.5	3.7	4080	2	US-08-446-345-35

ALIGNMENTS

RESULT 1
US-09-060-410-1
; Sequence 1, Application US/09060410
; Patent No. 6165461
; GENERAL INFORMATION:
; APPLICANT: Cobb, Melanie
; APPLICANT: Hutchinson, Michele
; APPLICANT: Chen, Zhu
; APPLICANT: Berman, Kevin
; TITLE OF INVENTION: TAO PROTEIN KINASES AND METHODS OF USE
; TITLE OF INVENTION: THEREFOR
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE DATA:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/060,410
; FILING DATE: 14-APR-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Makl, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 860098.421
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-5031
; INFORMATION FOR SEQ ID NO: 1:

Alignment Scores: 0.0451 Length: 2187
 Pred. No.: 145.00 Matches: 112
 Score: 36.04% Conservative: 88
 Percent Similarity: 20.18% Mismatches: 227
 Best Local Similarity: 4.67% Indels: 128
 Query Match: 24 Gaps: 20
 DB:

US-09-890-475-1 (1-609) x ABS51736 (1-2187)

Qy 34 LysIleValGluThrGluSerThrSerMetAspIleThrIleGlyGlnSerLysGlnPro 53
 Db 634 AAATCTGTTCAATAGAGAAAGAAAGATTGAT-----GAAATCTGAACAGAA 684
 Qy 54 GlnPheLeuLysSerIleAspGluLeuAlaIlePheSerValAlaValGluThrPheLys 73
 Db 685 AAATCTGGAATACATCAAGAAATAGTCTGCTTCAGATCAAGTGGAAAAATACAA 744
 Qy 74 ArgGlnPheAspAspLeuGlnLysHisIleGluSerIleGluAsnAlaIleAspSer--- 92
 Db 745 CTAGATATTGCCAGTTAGAGAAATTTGAAAGAGAGATGATGAATTTTAAAGCCTT 804
 Qy 93 -----LysLeuGluSerAsnGlyValValLeuAla----- 102
 Db 805 AAGCAGTCTCTTGAGGAGAAATATTGTATTATCTAAACAGTAGTAGAGATCTAAATGTG 864
 Qy 103 -----AlaArgAsnAsnAsnPheHisGln 110
 Db 865 AATGTCAGTGTCTGAAAGAAAGAAAGAACCATGTCACAGGAATAGAGAACACAC 924
 Qy 111 ProMetLeuSerProProArgAsnAsnValSerValGluThrValThrValSerGln 130
 Db 925 GAAATCTAAATCAGAGATGCAAAACTTA----- 954
 Qy 131 ProSerGlnGluLeuValProGluThrSerAsnLysProGluGlyGlyArgMetCysGlu 150
 Db 955 ---AAACAGAAAGTTATTCTTGACCAACAGGAACTG---GAAAGCTTCAACAAAAGAA 1008
 Qy 151 LeuMetCysSerLysGlyLeuArgLysIleTyrAlaAsnIleSerAspGlnAlaLys 170
 Db 1009 TTCAAAATGATTCACTCTCTGCAACAGAGAAAGAAATATTCTCGAGTCTTCATCAGAAG 1068
 Qy 171 Leu-----MetGluGluIleProSerAlaLeuLysLeuAlaValGluProAlaLys 187
 Db 1069 CTCTGTTCTTTTCAAGAGGAATGGTTAAAGAGAGAAATCTGTTGAGGAGAAATTAAG 1128
 Qy 188 PheValLeuAspCysIleGlyLysPheTyrLeu-----GlnGlyArgAlaPhe 204
 Db 1129 CAACACTGTGATGAGCTTGATAATATACAGCAAAAGAGGAAACAGCTGAAAGGCTGTC 1188
 Qy 205 ThrLysGluSerProMetSerAlaArgGlnValSerLeuLeuIleLeuGluSerPhe 224
 Db 1189 AAGCAATTGGAAGAGGAAGCAAAATCTAGAGCTGAAGAAATTAATCTCTAGAGAAAG 1248
 Qy 225 LeuLeuMetProAspArgGlyLysGlyLysValLysIleGluSerTrpIleLysAspGlu 244
 Db 1249 CTG-----AAAGGAAG-----GAGGCTGAA 1269
 Qy 245 AlaGluThrAlaAlaValAlaIlePheArgLysArgLeuMetThrGluGlyLeuAlaAla 264
 Db 1270 CTGGAGAAAGTAGTGCTGCT-----CATACCCAGCCACCCTGCTGCTTTG 1314
 Qy 265 AlaGluLysMetAspAla-----ArgGly 272
 Db 1315 CAGGAAAGATGATGAGATGTTGGCAAGCCTTGAGATGTTACTGCTCAATTTGAGGC 1374
 Qy 273 LeuLeuLeuValAlaCysPheGlyValProSerAsnPheArgSerThrAspLeuLeu 292
 Db 1375 TATAAAGCGTTTAAACAGCCAGTGAATAGAGATCTTAAGCTGGAGAACTCATCATACAG 1434
 Qy 293 AspLeuLeuArgMetSerGlySerAsn-----GluIleAlaGlyAla 306
 Db

Db 1435 GAAAGAGCGCCCAAGGCTGGGAAAAATGCAGAGATGTTTCAGCATCAGATTTTGGCAACT 1494
 Qy 307 LeuLysArgSerGlnPheLeuValProMetValSerGlyIleValGluSerSerIleLys 326
 Db 1495 GAGAGCTCAAAATCAAGATATGTAAAGATGCTTCTAGATCTGCAGACCAAGTCAGACTA 1554
 Qy 327 ArgGlyMetHisIleGluAlaLeuGluMetValTyrThrPheGlyMetGluAspLysPhe 346
 Db 1555 AAGGAAACAGAAATTAAGAAATCACAGTTCTTTCTTCAAAAAATAACTGAT----- 1608
 Qy 347 SerAlaAlaLeuValLeuThrSerPheLeuLysMetSerLysGluSerPheGluArgAla 366
 Db 1609 -----TTGCAGACCAACTCAAGCAACAGAGGAGAGACTTTAGAAAACAG 1653
 Qy 367 LysArgLysAlaGlnSerProLeuAlaPheLysGluAlaIleThrLysGlnLeuAlaVal 386
 Db 1654 CTGGAAGATGAAAGAGAGAAAGCTGAAAGAAATACACAGCAGAAATTA----- 1707
 Qy 387 LeuSerSerValMetGlnCysMetGluThrHisLysLeuAspProAlaLysGluLeuPro 406
 Db 1708 -----ACTGAAGAAATTAAC 1722
 Qy 407 GlyTyrGln---IleLysGluGlnIleValSerLeuGluLysAsp---ThrLeuGlnLeu 424
 Db 1723 AAGTGGCGTCTCTATGAAAGACTATATATAAAACAAACCTTTTCAGCTACACTA 1782
 Qy 425 AspLys---GluMetGluGluLysAlaArgSerLeuSerLeuMetGluGluAlaAlaLeu 443
 Db 1783 GATGCTTTTGAAGTAGAAAAACAGGCA-----TTGTTGAATGAACATGGTGCA 1830
 Qy 444 AlaLysArgMetTyrAsn-----GlnGln 451
 Db 1831 GCTCAGGAACAGCTAAATAAATAAGAGATTCATATGCTAAATTTATGGGTCAATCAAT 1890
 Qy 452 IleLysArgProArgLeuSerProMetGluMetProProValThrSerSerSerTyrSer 471
 Db 1891 TTCAACAAATAATCAAGCATGTTCTGGAAGTTGAAGATGAATAAGTAGCAACTCAATCG 1950
 Qy 472 ProIleTyrArgAspArgSerPheProSerGlnArgAspAspGlnAspGluIleSer 491
 Db 1951 GAAGTATCAAAACTCCGCTGTCAGCTTCTGCTAAAAAAACAAAGTGAGACAAAACCTCAA 2010
 Qy 492 AlaLeuValSerSerTyrLeuGly-----ProSerThrSerPheProHis 506
 Db 2011 GAGGAATTGAATAAAGTTCTAGGTATCAACACATTGATCTTCAAGGCTTTTTCATCAT 2070
 Qy 507 ArgSerArgArgSerProGluTyrMetValProLeuProHisGly 521
 Db 2071 GAAAGTAAAGAAATTTTGGCCCTGAAGACCCCATTAAGAAAGGC 2115

Search completed: August 25, 2003, 10:39:46
 Job time : 547 secs

Db 1459 CAGAGGCTGCCATCTG-----CAGACTGAGGTGGATGCT 1494
QY 116 ProArgAsnAsnValSerValGluThrThrValThrValSerGlnProSerGlnGluIle 135
Db 1495 CTCGGATTGCGTTTGAAGAGAGAAACCATG--TTGAATAAAAGACAAACCAATT 1551
QY 136 ValProGluThrSerAsnLysProGluGlyArgMetCysGluLeuMetCysSerLys 155
Db 1552 CAGGATATGGCTGAAGAGAG-----GGGACACAGCTGGAGAGATACATGACCTCAAG 1605
QY 156 GlyLeuArgLysTyrIleTyrAlaAsnLysSerAspGlnAlaLysLeuMetGluGluIle 175
Db 1606 GACATGTTGGATGTCAAGGAGCGGAAGTTAATGTTCTTCAGAGAAGATTGAATAATCTT 1665
QY 176 ProSerAlaLeuLysLeuAlaLysGluProAlaLysPheValLeuAspCysIleGlyLys 195
Db 1666 CAAGAGCAGCTTAGA-----1680
QY 196 PheTyrLeuGlnGlyArgAlaPheThrLysGlySerProMetSerSerAlaArgGln 215
Db 1681 -----CACAGGAAGAGCAGATGACGACTTGAAGAA 1713
QY 216 -----ValSerLeuLeuIleLeuGlu 222
Db 1714 CGGTCAATCTCCAGCGCTGCACCAACACAGCTGACCTGCTTGACAACTTTGGAG 1773
QY 223 SerPheLeuLeuMetProAspArgGlyLysGlyLysValIleGluSerTyrPheLys 242
Db 1774 GAGGCCCTTCAGAGAGAGCGGCAATTGAACGCTTAAAG---GAGCAGAGGGACAGA 1830
QY 243 AspGluAlaGluThrAlaAlaValAlaTyrPheArgLysArgLeuMetThrGluGlyLeu 262
Db 1831 GATGAGCGAGAG-----1842
QY 263 AlaAlaAlaGluLysMetAspAlaArgGlyLeuLeuLeuValAlaCysPheGlyVal 282
Db 1843 ----AAGCAGAGGAATTCAT-----1860
QY 283 ProSerAsnPheArgSerThrAspLeuLeuAsp-----LeuIleArg 296
Db 1861 -----AATCAAAAAA---GATCTTAAGACTTGAAGAAAAAGTCAGCTGTTGCAA 1911
QY 297 MetSerGlySerAsnGluIleAlaGlyAlaLeuLysArgSerGlnPheLeuValProMet 316
Db 1912 GCGACCTTTCAGAGAAAGAGCTTCATTTTGGAT-----CTGAAGAGCAT 1959
QY 317 ValSerGlyIleValGluSerSerIleLysArgGlyMetHisIleGluAlaLeuGluMet 336
Db 1960 GCTTCTCTCGCATCTCCAGACTGAAAAGGACTCACGGCTTAAGACACTAGAGATT 2019
QY 337 ValTyrThrPheGlyMetGluAspLysPheSerAlaAlaLeuValLeuThrSerPheLeu 356
Db 2020 -----GCTTTGGAGCAGAGAAGGAGGAGTGTCCTGAAAATGGAATCACAATTG 2067
QY 357 Lys-----MetSerLysGluSerPheGluArgAla 366
Db 2068 AAAAGGCACATGAGCGCAGCATTTGAACCCAGACCGAGTCAGAGATGATGACCGAATA 2127
QY 367 LysArgLysAlaGlnSerProLeuAlaPheLysGluAlaThrLysGlnLeuAlaVal 386
Db 2128 CAGCAGCTTGGAGAGAGAGATCACCAGGTACAAAGATGAATCTAGCAGGCCAGGCGAGAA 2187
QY 387 LeuSerSerValMetGluCysMet-----GluThrHisLysLeuAspProAlaLys 403
Db 2188 GTTATCGACTTTAGAAAATCTTGAGAGAGGTGAAATGAGAAAGATGAAAGATGAATG 2247
QY 404 Glu-----LeuProGlyTyrGlnLysGluGln-----Ile 414
Db 2248 AAGATAGCTGAGTTGCAAGTCTCACTCAAGCAAGTGAAGAACACAGATAAGAAGGTA 2307
QY 415 ValSerLeuGluLysAspThrLeuGlnLeuAspLysGluMetGluGluLysAlaArgSer 434
Db 2308 GCAAAATCTGAAG-----CACAGGAACAGGTGGGAAAAAAGAGAGAT 2349

QY 435 LeuSerLeuMetGluGluAla 441
Db 2350 GCACAAATGTTAGAGGAGGCG 2370

RESULT 15

ABSS1736
ID ABSS1736 standard; cDNA; 2187 BP.
XX ABSS1736;
XX 21-OCT-2002 (first entry)
XX Human novel polynucleotide #9.
DE
XX Human; gene; as; trauma; viral infection; parasitic infection; addiction;
KW bacterial infection; Alzheimer's disease; Huntington's disease; anxiety;
KW Parkinson's disease; behavioural disorder; pain; hair growth disease;
KW alopecia; pigmentation disorder; inflammatory disorder; arthritis; AIDS;
KW inflammatory bowel disease; Crohn's disease; cancer; adenocarcinoma;
KW acquired immunodeficiency syndrome; colon; asthma; hypertension; obesity;
KW autoimmune disease; diabetes; graft versus host disease; ulcer; bulimia;
KW anorexia; dementia; gene therapy.
XX
XX Homo sapiens.
XX WO200257452-A2.
XX 25-JUL-2002.
XX 17-DEC-2001; 2001WO-US49122.
XX 15-DEC-2000; 2000US-256025P.
PR 30-JAN-2001; 2001US-265163P.
PR 02-MAR-2001; 2001US-272929P.
PR 09-MAR-2001; 2001US-274864P.
PR 16-MAR-2001; 2001US-276888P.
PR 22-MAR-2001; 2001US-277880P.
PR 25-APR-2001; 2001US-286409P.
PR 31-JUL-2001; 2001US-309246P.
PR 29-AUG-2001; 2001US-315600P.
XX (CURA-) CURAGEN CORP.
PA Shimkets RA, Colman SD, Spytek KA, Ballinger RA, Guo X;
PI Tchernov VT, Shenoy SC, Li L, Ellerman KE, Zerhusen BD;
PI Patturajan M, Casman SJ, Boldog F, Gusev VV, Burgess CE;
PI Edinger S, Gangolli EA, Malyankar UM, Gunther E, Smithson G;
PI Millet I, Gerlach VL;
XX WPI; 2002-590743/63.
DR P-PSDB; ABG70293.
XX Novel polypeptide, designated NOVX for treating or preventing disorders
or symptoms e.g. trauma, Alzheimer's disease, cancers, acquired
immunodeficiency syndrome, asthma and rheumatoid arthritis -
XX Claim 9; Page 47; 252pp; English.
XX The invention relates to human novel polynucleotides and polypeptides.
The sequences are useful for the treatment, prevention and diagnosis of
disorders such as trauma, viral/parasitic/bacterial infections,
Alzheimer's disease, Huntington's disease, Parkinson's disease,
behavioural disorders, anxiety, addiction, pain, hair growth diseases,
alopecia, pigmentation disorder, inflammatory disorders such as osteo-
arthritis, inflammatory bowel disease, Crohn's disease,
acquired immunodeficiency syndrome (AIDS), cancers such as colon cancer
and adenocarcinoma, asthma, hypertension, autoimmune disease, diabetes,
obesity, graft versus host disease, ulcer, bulimia, anorexia and
dementia. Sequences ABSS1728-ABSS1748 represent human novel
polynucleotides of the invention.
XX Sequence 2187 BP; 872 A; 354 C; 461 G; 500 T; 0 other;


```

DT 28-JUN-2002 (first entry)
XX
DE Novel human coding sequence SEQ ID NO: 150.
XX
KW Human; antianaemic; vulnery; antiinflammatory; immunomodulator;
KW antiferility; cerebrotective; cytotatic; rheumatic; gene therapy;
KW neuroprotective; antiparkinsonian; protein therapy; EST;
KW expressed sequence tag; gene; ss.
XX
DS Homo sapiens.
XX
FN WO200222660-A2.
XX
PD 21-MAR-2002.
XX
PF 10-SEP-2001; 2001WO-US26015.
XX
PR 11-SEP-2000; 2000US-0659671.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;
PI Xue AJ, Yang Y, Wehrman T, Drmanac RT;
XX
DR WPI; 2002-292408/33.
DR P-PSDB; ABB97326.
XX
PT An isolated polynucleotide for treating diseases associated with its
XX encoded polypeptide such as cancer and multiple sclerosis -
XX
PS Claim 1; SEQ ID NO 150; 509pp; English.
XX
CC The present invention provides the protein and coding sequences of 444
CC novel human proteins. These were isolated from expressed sequences tags
CC (ESTs). They can be used to stimulate cell growth, to regulate
CC haematopoiesis e.g. to treat aplastic anaemia, to help tissue regrowth
CC e.g. in burn treatment, to regulate the immune system e.g. to treat
CC multiple sclerosis, to regulate activin or inhibin e.g. to treat
CC infertility, to regulate haemostasis or thrombolysis e.g. to treat
CC stroke and cancer, to screen for drugs, to treat inflammatory conditions
CC e.g. rheumatoid arthritis, and to treat nervous system disorders e.g.
CC Parkinson's disease. The present sequence is a coding sequence of the
CC invention.
XX
SQ Sequence 4972 BP; 1579 A; 1001 C; 1127 G; 1263 T; 2 other;

Alignment Scores:
Pred. No.: 0.0719 Length: 4972
Score: 148.50 Matches: 150
Percent Similarity: 34.98% Conservative: 98
Best Local Similarity: 21.10% Mismatches: 227
Query Match: 4.78% Indels: 237
JB: 24 Gaps: 35

US-09-890-475-1 (1-609) x ABB59739 (1-4972)
2Y 3 AsnTyr-----ProThrValAlaAlaGlnProThrThrAlaAsn 17
Db 1443 AATTACAGAGAAGGGAGATCTTAGAACAGAGCATCAATCTCCA-----1496
2Y 18 ProLeuGlnArgHisGlnSerGluGlnArgArgGluLeuProHisLeValGlu 37
Db 1497 CCCCAAGTATCTGTCACAAATCACACTATGTAATCGAGAACACTTTGCTACTATACGG 1556
2Y 38 ThrGluSerThrSerMetAspIleThrIleGlnSerIleGlnSerIleGlnSerIleGlnSer 57
Db 1557 ACAGCATCACTG-----CTTACGAGGCAATATGCAAGAACATGAG 1595
2Y 58 SerIleAspGluLeuAlaAlaPheSerValAlaValGluThrPheLysArgGlnPheAsp 77
Db 1596 CAGGACTCTGAGCTTAGAGAACAAATGCTGCTATAGCGAATGAGCGGACACAT---1652
2Y 78 AspLeuGlnLysHisIleGluSerIleGluAsnAlaIleAspSerLysLeuGluSerAsn 97

```

1653 Db -----CAAAGCAACTGATGACTCTGGAAAACACAGCTAAAGCTGAGATGGATGAACAT 1706
 98 QY GlyValValLeuAlaAlaArgAsnAsnAsnPheHisGlnProMetLeuSerProProArg 117
 1707 Db CGCTCAGATTA-----GACAAAAGATCTTGAACACTCAGCGT 1742
 118 QY AsnAsnValSerValGluThrThrValThrValSerGlnProSerGlnGluLeuValPro 137
 1743 Db AACAAATTTCTGCAGAAATGGAGAACTTATCAAGAAACACACAGGCTCTATGAGAAA 1802
 138 QY GluThr-----SerAsnLysProGluGlyGlyArgMetCysGluLeuMetCysSer 154
 1803 Db GAGCTAAAGTGATGTCATGAA-----GAG 1829
 155 QY LysGlyLeuArgLysTyrIleTyrAlaAsnIleSerAspGlnAlaLysLeuMetGluGlu 174
 1830 Db AAAAAATTCAGCAACATATT-----CAGCCCCACAGAGAAAGAA 1871
 175 QY IleProSerAlaLeu-----LysLeuAlaLysGluProAlaLys 187
 1872 Db CTGAATAGTTTCTCGAGTCCAGAAAAGAGAGTATAAACTTCGAAAAGAGAGCTTAA 1931
 188 QY PheValLeuAspCysLysGlyLysPheTyrLeuGlnGlyArg-----AlaPhe 204
 1932 Db GAGGAGCTAAAT---GAAAACACAGAG-TACCCCCAAAAAGAAACACAGAGTGCTTTC 1987
 205 QY ThrLysGluSer-----PrometSerSer 212
 1988 Db AAGCAGAGAGAGAAATATACAGCATTTCCAGCAGAGAGAGAGTAACTCTTCGAGC 2047
 213 QY AlaArgGlnValSerLeuLeuLeuGluSerPheLeuLeuMetProAspArgGlyLys 232
 2048 Db TCMAAGACAATA-----CCTAGAGCTGGAATG 2074
 233 QY GlyLysValLysIleGluSerTrpIleLysAspGluAlaGluThrAlaAlaValAlaTsp 252
 2075 Db CCGTCGGTTCAAGAGAAGAT-----GTTACTTGG 2104
 253 QY ArgLysArgLeuMetThr-----GluGlyGlyLeuAlaAlaGluLysMetAsp 269
 2105 Db CGGTCAATACTTAGACGAGACCTTGTCCAGGAGGAGT-----AAACAAAAGACAGAC 2158
 270 QY AlaArgGlyLeuLeuLeuValAlaCysPheGlyValProSer-----284
 2159 Db TCAGAGGAGCTTAGA-----GCATGCCATGCTACTCCGACAGCATGAATCTATGCA 2209
 285 QY -----AsnPheArgSerThrAspLeuLeu-AspLeuLeuAeMgMe 297
 2210 Db AGAAGTGGAGTTCGGCCACCTCAACACAATTCAGAAAGATCGCGTGTGAGTTGATCAGATT 2269
 297 QY tSerGlySerAsnGluLeuAlaGlyAlaLeuLysArgSerGlnPheLeuValProMetVa 317
 2270 Db ACAGCATCAAACTGAGCTCACTAACCCAGCTGGAATATATAAGCGGAAGA-----2318
 317 QY LserGlyLeuValGluSerSerIleLysArgGlyMetHisIleGluAlaLeuMetVa 337
 2319 Db -----GACGAGAACTAGACGA---AAGCATGTC-----2345
 337 QY lTyThrPheGlyMetGluAspLysPheSerAlaAlaLeuValLeuThrSerPheLeuLys 357
 2346 Db -----ATGAGAGTTCGACACAGCCCTAAGAGTTCGAAAGTCTTAAAGAACTCCA 2392
 357 QY smetSerLysGluSerPheGluArgAlaLysArgLysAlaGlnSerProLeuAla-----375
 2393 Db AATAAAAACAGTTTCAGGATACCTGCAAAATCCAAACACAGACAGTACAAAGCATTANG 2452
 376 QY -----Phely 377
 2453 Db AATCACCTGCTGAGAGTACCAAAAGAGTGAAGCAAAAGTGTCTTCTGAAACGGGTCAA 2512
 377 QY sGluAlaAlaThrLysGlnLeuAlaValLeuSer-----SerValMetCl 392

Qy 18 ProLeuLeuGlnArgHisGlnSerGluGlnArgArgGluLeuProLysLeuValGlu 37
Db 1419 CCCAAGTATCTCGTCAACAAACACACTATCGTAATCGAGAACACTTTGCTACTATACGG 1478
Qy 38 ThrGluSerThrSerMetAspIleThrIleGlyGlnSerLysGlnProGlnPheLeuLys 57
Db 1479 ACAGCATCTG-----GTTACGAGCGCAATGCAAGAACATGAG 1517
Qy 58 SerIleAspGluLeuAlaPheSerValAlaValGluThrPheLysArgGlnPheAsp 77
Db 1518 CAGGACTCTGAGCTTAGAGAACAAATGCTGCTGATATAGCGAATGAGCGCAACAT--- 1574
Qy 78 AspLeuGlnLysHisIleGluSerIleGluAsnAlaIleAspSerLysLeuGlnSerAsn 97
Db 1575 -----CAAAGCACTGATGCTCTGGAACAAAGCTAAGGCTGAGATGATGAACAT 1628
Qy 98 GlyValValLeuAlaAlaArgAsnAsnPheHisGlnProMetLeuSerProProArg 117
Db 1629 CGCCTCAGATTA-----GACAAAGATCTTGAAACTCAGCGT 1664
Qy 118 AsnAsnValSerValGluThrThrValThrValSerGlnProSerGlnGluIleValPro 137
Db 1665 AACAAATTTCTGCTGAGAAATGAGAAACTTATCAAGAAACACAGGCTGCTATGGAGAA 1724
Qy 138 GluThr-----SerAsnLysProGluGlyGlyArgMetCysGluLeuMetCysSer 154
Db 1725 GAGGCTAAAGTGATGTCATGAA-----GAG 1751
Qy 155 LysGlyLeuArgLysIleValAlaAsnIleSerAspGlnAlaLysLeuMetGluGlu 174
Db 1752 AAAAAATTCAGCAACATATT-----CAGGCCCAACAGAGAAAGAA 1793
Qy 175 IleProSerAlaLeu-----LysLeuAlaLysGluProAlaLys 187
Db 1794 CTGAATAGTTTCTCGAGTCCCAAGAAAGAGAGATATAAACTTCGAAAGAGCAGCTTAA 1853
Qy 188 PheValLeuAspCysIleGlyLysPheThrLeuGlnGlyArg-----AlaPhe 204
Db 1854 GAGGAGCTAAAT---GAAACCCAGAG-TACCCCAAAAGAAACACAGGAGTGCTTTC 1909
Qy 205 ThrLysGluSer-----ProMetSerSer 212
Db 1910 AAGCAGAAGGAGAAATATACAGCATTTCCAAAGCAGAAGAAAGAGCTTCTTCTCGAGC 1969
Qy 213 AlaArgGlnValSerLeuLeuIleLeuGluSerPheLeuLeuMetProAspArgGlyLys 232
Db 1970 TCAAAGACAATA-----CCTAGAGCTGGAATG 1996
Qy 233 GlyLysValLysIleGluSerTrpIleLysAspGluAlaGluThrAlaAlaValAlaTrp 252
Db 1997 CCCTCGCTTCAAGAGAAGAAAT-----GTTACTTGG 2026
Qy 253 ArgLysArgLeuMetThr-----GluGlyGlyLeuAlaAlaAlaGluLysMetAsp 269
Db 2027 CGCTCATAACTTAGACGAGCAGCACCTTGTGCGGAGGAGTT-----AAACAAAGACAGAC 2080
Qy 270 AlaArgGlyLeuLeuLeuValAlaCysPheGlyValProSer----- 284
Db 2081 TCAGAGAGACTTAGA-----GCAATGCATGCTACTCCGACAGCATGATCTATGCA 2131
Qy 285 -----AsnPheArgSerThrAspLeuLeu-AspIleuIleArgMe 297
Db 2132 AGAACTGGAGTTCCGCCCACTCAACACAATTCAGAGATCGCTGTGAGTTGATCAGATT 2191
Qy 297 tSerGlySerAsnGluIleAlaGlyAlaLysArgSerGlnPheLeuValProMetVa 317
Db 2192 ACAGATCAAACTGAGTCACTACCACTGGAATATAATAAGCGAAGA----- 2240
Qy 317 lserGlyIleValGluSerSerIleLysArgGlyMetHisIleGluAlaLeuGluMetVa 337
Db 2241 -----GAACGAGAAGCTAAGACGA---AAGCATGTC----- 2267
Qy 337 lTyrThrPheGlyMetGluAspLysPheSerAlaAlaLeuValLeuThrSerPheLeuLys 357

Db 2268 -----ATGGAAGTTTCGACCAACAGCCTTAAGAGTTTGAAGTCTAAAGACTCCA 2314
Qy 357 sMetSerLysGluSerPheGluArgAlaLysArgLysAlaGlnSerProLeuAla----- 375
Db 2315 AATAAAAAGCAGTTTTCAGGATACCTCGCAAAANTCCAAACACAGACAGTACAAAGCATTAAG 2374
Qy 376 -----PheLys 377
Db 2375 AAATCACCTGCTGGAGACTACACCAAAAGAGTGACACAAAGCTGTTCTCGAAACGGCTCAA 2434
Qy 377 sGluAlaAlaThrLysGlnLeuAlaValLeuSer-----SerValMetG1 392
Db 2435 GGAGAAACAGACCCGGAATAGCTATCTTGGCTGAGCAGTATGATCAGACCATTAATGA 2494
Qy 392 nCysMetGluThrHis-----LysLeuAspProAlaLysGluLeuProGlyTrpGlnI1 410
Db 2495 AATGCTCTCCACACAAAGCCCTGCGTTTGGATGAAGCACAGGAAGCAGAGTCCAGGTTTT 2554
Qy 410 eLysGluGlnIle-----ValSerLe 417
Db 2555 GAAGATGCGAGCTGCAGCAGGAACCTGGAGCTGTTGAATCGGTATCAGAGCAAAATCAAGAT 2614
Qy 417 uGluLysAspThrLeuGlnLeuAspLysGluMetGluGluLysAlaArgSerLeuSerLe 437
Db 2615 GCAAGCTGAGGCA---CAACATGATCGAGAGCTTCGCGAGCTTGAACAGAGGGTCTCCCT 2671
Qy 437 u-----MetGluGluAlaAlaLeuAlaLysArgMetTyrAs 449
Db 2672 CGGAGGCGCACTCTTAGAACAAAGATTGAAGAGAGATGTTGGCT-----TTGCAGAA 2725
Qy 449 nGlnGlnIleLysArgProArgLeuSerProMetGluMetProValThrSerSerSe 469
Db 2726 TGAGCCACAGAACGAAATACGA----- 2747
Qy 469 rTyrSerProIleTyrArgAspArgSerPheProSerGlnArgAspAspGlnAspG1 489
Db 2748 -----AGCCTGTTGGAACGTCACAGCCAGAGA 2773
Qy 489 uIleSerAlaLeuValSerSerTyrLeuGlyProSerThrSerPheProHisArgSerAr 509
Db 2774 GATTGAAGCTTTTGACTCTGAAGCATGAGACTAGGTTTATGTAATATGTCCTTTCTAA 2833
Qy 509 gArgSerProGluTyrMetValProLeuProHisGlyLeuGlyArgSerValTyrAl 529
Db 2834 TCTCTCCCTGAG-----GC 2848
Qy 529 aTyrGluHisLeuAlaProAsn-----SerTyrSerProGlyHisGly----- 543
Db 2849 ATTCAAGCAGACAGTACCCCGAGAGCTTGGTGGTTCACACACACCTCTAGGGGTCCAGG 2908
Qy 544 ---HisArgLeuHisArgGlnTyrSerProSerLeuValHisGlyGlnArgHisProLe 562
Db 2909 ACCTCACTGGGTTCATCCCATGGTGGCCCAACCAAGCTTGGGC-----CATCCAAT 2962
Qy 562 uGlnTyrSerPro-----ProIleHisGlyGlnGlnLeuProTyrG1 577
Db 2963 GCAAGGTGGACCCCAAGCCATCGGCTCACCCCTTCAAGGGCAATGCAAGGGTACCTCCAGG 3022
Qy 577 yIleGlnArgValTyrArgHisSerProSerGluGluArgTyrLeu-----GlyLe 594
Db 3023 TAGCAGTATGGGAGTCCGAATAGCCCCCAGGCTCTGAGGCGGACAGCTTCTGGGGGACG 3082
Qy 594 uSerAsnGlnArgSerProArgSerAsnSer 604
Db 3083 GACAGACGAGGATGAGCAGAGCAGAGT 3113
RESULT 13
AENS9739 standard; cDNA; 4972 BP.
ID ABNS9739
XX
AC ABNS9739;
XX

```
QY 317 lserGlylleValGluSerSerlleLysArgGlyMetHieLleuAlaLeuGluMetVa 337
Db 1445 -----GAACGAGAACTAAAGCGA---AAGCATGTC----- 1471
QY 337 lTyrThrPheGlyMetGluAplLysPheSerAlaAlaLeuValLeuThrSerPheLeuLy 357
Db 1472 -----ATGGAAGTTCGACACAGCCCTAAGAGTTTGAAGTCTAAAGAACTCCA 1518
QY 357 smetSerLysGluSerPheGluArgAlaLysArgLysAlaGlnSerProLeuAla----- 375
Db 1519 AATAAAAAGCAGTTTCAGGATACCTCGCAAAATCCAAACAGCAGACATCAAGCAATTAAG 1578
QY 376 -----PheLy 377
Db 1579 AAATCACCTGTGGAGACTACACAAAGAGTGAGCACAAAGCTTCTGAAACGGCTCAA 1638
QY 377 sGluAlaAlaThrLysGlnLeuAlaValLeuSer-----SerValMetGl 392
Db 1639 GGAGGAACAGACCCCGAAATTAGCTATCTTGGCTGAGCAGTATGATCACAGCATTAAATGA 1698
QY 392 nCysMetGluThrHis-----LysLeuAspProAlaLysGluLeuProGlyTrpGlnIl 410
Db 1699 ATGTCTCTCCACACAGACCCCTGCGTTTGGATGACACAGACAGAGTCCCGGTTT 1758
QY 410 elysGluGlnIle-----ValSerLe 417
Db 1759 GAAGATGCAGCTGCAGCAGGAAGTGGAGCTGTGAATCGCTATCAGACGAAAATCAAGAT 1818
QY 417 uclulLysAspThrLeuGlnLeuAplLysGluMetGluGluLysAlaArgSerLeuSerLe 437
Db 1819 GCAAGCTGAGGCA---CAACATGATCGAGAGCTTCGGAGCTTGAACAGAGGTCTCCCT 1875
QY 437 u-----MetGluGluAlaAlaLeuAlaLysArgMetTyrAs 449
Db 1876 CCGGAGGGCACTCTTAGAACAAAAGATTGAAGAAGAGATGTGGCT-----TTGCAGAA 1929
QY 449 nGlnGlnLleLysArgProArgLeuSerProMetGluMetProProValThrSerSerSe 469
Db 1930 TGAGCGCAGAGCAAGATACGA----- 1951
QY 469 rTyrSerProIleTyrArgAspArgSerPheProSerGlnArgAspAspGlnAspGl 489
Db 1952 -----AGCTGTGGAACGTCAAGCCAGAGA 1977
QY 489 ulleSerAlaLeuValSerSerTyrLeuGlyProSerThrSerPheProHisArgSerAr 509
Db 1978 GATTGAAGCTTTTGACTCTGAAGCATGAGACTAGGTTTAGTAAATATGTCCTTTCTAA 2037
QY 509 gArgSerProGluTyrMetValProLeuProHisGlyGlyLeuGlyArgSerValTyrAl 529
Db 2038 TCTCTCCCTGAG-----GC 2052
QY 529 aTyrGluHisLeuAlaProAsn-----SerTyrSerProGlyHisGly----- 543
Db 2053 ATTACGCCACAGCTACCCGGGAGCTCTGGTTGGTCAACACACCTACTCGGGGTCCAGG 2112
QY 544 ----HisArgLeuHisArgGlnTyrSerProSerLeuValHisGlyGlnArgHisProLe 562
Db 2113 ACCTCACTGGGTATCCATGGTGGGCCACACAGCTTGGGCC-----CATCCAAAT 2166
QY 562 uGlnTyrSerPro-----ProIleHisGlyGlnGlnLeuProTyrGl 577
Db 2167 GCAAGGTGACCCAGCCATCGGGTCAACCTTCAGGGCCAAATGCAAGGGGTACTCGAGG 2226
QY 577 yIleGlnArgValTyrArgHisSerProSerGluGluArgTyrLeu-----GlyLe 594
Db 2227 TAGCATGATGGAGTCCGCAATAGCCCCAGGCTCTGAGCGGACAGCTTCTGGGGGACG 2286
QY 594 uSerAsnGlnArgSerProArgSerAsnSer 604
Db 2287 GACGAGCAGGGCATGAGCAGAGGACCAAGT 2317
```

```
RESULT 12
ACC50205
ID ACC50205 standard; cDNA; 4620 BP.
XX
AC ACC50205;
XX
DT 12-JUN-2003 (first entry)
XX
DE Breast cancer associated cDNA sequence SEQ ID NO:253.
XX
KW Human; breast cancer; cytostatic; gene therapy; gene; ss.
XX
OS Homo sapiens.
XX
PN WO2003004989-A2.
XX
PD 16-JAN-2003.
XX
PF 21-JUN-2002; 2002WO-US19669.
XX
PR 21-JUN-2001; 2001US-299887P.
PR 27-JUN-2001; 2001US-301572P.
PR 18-JUL-2001; 2001US-306501P.
PR 25-SEP-2001; 2001US-325002P.
PR 05-MAR-2002; 2002US-362585P.
PR 14-MAY-2002; 2002US-380391P.
XX
PA (MILL-) MILLENIUM PHARM INC.
XX
PI Lillie J, Gannavaru M, Glatt K, Hoersh S, Kamatkar S, Mertens M;
PI Monahan JE, Myer V, Wang Y, Xu Y, Zhao X, Meyers RE, Bast RC;
PI Hortobagyi GN, Pusztai L, Meric F, Sahin A, Mills GB;
XX
WPI; 2003-210381/20.
P-PSDB; ABR47509.
XX
PT Breast cancer diagnosis or treatment by comparing the level of
PT expression of a marker in a patient sample with that in the control
XX non-breast cancer sample
PS Claim 1; SEQ ID 253; 128pp; English.
XX
CC The present invention describes a method for assessing whether a patient
CC is afflicted with breast cancer. The method comprises comparing the level
CC of expression of a marker (gene/polypeptide see ACC50076 to ACC50334 and
CC ABR47386 to ABR47632) in a patient sample and the normal level of
CC expression of the marker in a control non-breast cancer sample, where a
CC significant increase in the level of expression of the marker in the
CC patient sample and the normal level is an indication that the patient is
CC afflicted with breast cancer. The breast cancer associated sequences
CC from the present invention have cytostatic activities and can be used in
CC gene therapy. The method is useful for diagnosing and treating breast
CC cancer.
CC N.B. The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 4620 BP; 1490 A; 942 C; 1039 G; 1149 T; 0 other;
Alignment Scores:
Pred. No.: 0.065 Length: 4620
Score: 148.50 Matches: 150
Percent Similarity: 34.88% Conservative: 98
Best Local Similarity: 21.10% Mismatches: 227
Query Match: 4.78% Indels: 237
DB: 25 Gaps: 35
US-09-890-475-1 (1-609) x ACC50205 (1-4620)
QY 3 AsnTyr-----ProProThrValAlaAlaGlnProThrThrAlaAsn 17
Db 1365 AATTACAGAGAAGAGGAGATCCTAGAACAGAGCATCAGATCCACAATCTCCA----- 1418
```


1333 AATTACCAAGAGAGAGATCTCTAGAACAGAGCATCAGCTCCACAGTCTCCA----- 1386
18 ProLeuLeuGlnArgHisGlnSerGluGlnArgArgGluLeuProLysLeuValGlu 37
1387 CCTCAAGTGTCTCGTCAAAATCAATCATATTATCGTAATAGAGAACACTTTTCAACTATACGA 1446
38 ThrGluSerThrSerMetAspIleThrIleGlyGlnSerLysGlnProGlnPheLeuLys 57
1447 ACAGCATCACTG-----GTTACAGACAGATGCAAGAACATGAG 1485
58 SerIleAspGluLeuAlaAlaPheSerValAlaValGluThrPheLysArgGlnPheAsp 77
1486 CAGGACTCTGAACCTAGAGAACAGATCTCTGGTTATAAGCGGATGAGCGACAGCAT--- 1542
78 AspLeuGlnLysHisIleGluSerIleGluAsnAlaIleAspSerLysLeuGluSerAsn 97
1543 -----CAGAAGCAGCTGATGACTCTCTGGAATAAATACTGAGGCGAGAAATGAGCAGACAT 1596
98 GlyVal-----ValLeuAlaAlaArgAsnAsnAsnPheHisGlnProMetLeu 113
1597 CGGCTCAGATTAGACAAAGATCTTTGAAACTCAGCGCAACAATTTTCGCTGCAGAAATG--- 1653
114 SerProProArgAsnAsnValSerValGluThrThrValThrValSerGlnProSerGln 133
1654 -----GAGAACTATTATGAGAAACACCAAGCTTCTATG 1686
134 GluIleValProGluThrSerAsnLysProGluGlyArgMetCysGluLeuMetCys 153
1687 GAAAAAGAGGCTAAAGTGTAGGCGCAACAGAG----- 1719
154 SerLysGlyLeuArgLysTyIleTyAlaAsnIleSerAspGlnAlaLysLeuMetGlu 173
1720 ---AAAAAATCCAAACACACTT-----CAGGCTCAACAGAGAATA 1758
174 GluIleProSerAlaLeuLys-LeuAlaLysGluProAlaLysPheValLeuAspCys1 193
1759 GAACCTGAATAGCTTTTGGAGTCTCAAAAGAGAGATATAAATCTCGAAAGAGCAGCTT 1818
193 eGlyLysPheTyLeuGlnGlyArgAlaPheThrLysGluSer-----ProMetSe 211
1819 AAGGAGGAGCTGAATGAACACAGCAGACACCTTAAAGAAAGAGAGGAGGATGGCTTTCA 1878
211 rSerAlaArgGlnValSerLeuLeuIleLeuGluSerPheLeuLeuMetProAspArg1 231
1879 AAGCAGAGGAGATATTCAACATTTTCAGCAGAGAGAGAGCTAATCTCTTCGACGT 1938
231 YLysGlyLysVal-----LysIleGluSerTrpIleLysAspGluAlaGluTh 247
1939 CAAGGCGATATCTAGAGCTAGAATGTCGCTGCTCA-----AAAGAGAATGTTTACTT 1992
247 rAlaAlaValAlaIleAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 267
1993 GGTCCGCAATACTTGAACAGAGCTTG-----TCAGGAGAGGTT-----AAACAAA 2040
267 yMetAspAlaArgGlyLeuLeuLeu-----LeuValAlaC 279
2041 AGGCAGACTCAGAGAGCTATGAGACATGCAATGTTACTGCGCAGCAGATGAATCCATGCAA 2100
279 yPheGlyVal---ProSerAsnPheArgSerThrAspLeuLeu-AspLeuIleArgMet 297
2101 GAACCTGGAGTTTCGCCACTCAACACTATTGAGAGATGGCTGTGAGTTGATCAGACTG 2160
298 SerGlySerAsnGluIleAlaGlyAlaLeuLysArgSerGlnPheLeuValProMetVal 317
2161 CAACATCAAACTGAGCTTACTTAACCCAGCTGGAATACAAATAAGAGAGG----- 2208
318 SerGlyIleValGluSerSerIleLysArgGlyMetHisIleGluAlaLeuGluMetVal 337
2209 -----GAACGGGAAGCTAAGACG---AAACATGTC----- 2235
338 TyrThrPheGlyMetGluAspLysPheSerAlaAlaLeuValLeuThrSerPheLeuLys 357
2236 -----ATGGAAGTTCGACAGCAGCCTTAAGAGTTTGAAGTTCTAAGAGACTCCAA 2283

QY 358 MetSerLysGluSerPheGluArgAlaLysArgLysAlaGlnSerProLeuAla----- 375
DB 2284 ATAAAAAGCAGTTTCAGGATACCTTCGAAAATTCAAACAGACAGTACAAAGCATTAAGG 2343
QY 376 -----PheLys 377
DB 2344 AATCACCCTACTGGAGACTACACAAAGAGTGAACAAAGCTGTTCTGAAAAGACTCAAG 2403
QY 378 GluAlaAlaThrLysGlnLeuAlaValLeuSer-----SerValMetGln 392
DB 2404 GAGAAACAGACTCGGAAGTTAGCCATCTTGGCTGAGCAGATGATCATAGCATTAATGAA 2463
QY 393 CysMetGluThrHis-----LysLeuAspProAlaLysGlu----- 404
DB 2464 ATGCTCTCCACACAAGCTCTGCGTTGGATGAAGCAGACAGGAAGCAGAAATGCCAGGTTTG 2523
QY 405 -----LeuProGlyTrpGlnIleLys----- 411
DB 2524 AAGATGCAAGTACAGCAGGAACCTGGAGCTTGTGAATGCATATCAGAGCAAAATCAAGATG 2583
QY 412 -----GluGlnIleValSerLeu----- 417
DB 2584 CAGGCTGAGGCCCAACATGATCGAGAGCTTCGAGAGCTGGAACAAAGGGTCTCCCTTCGG 2643
QY 418 -----GluLysAspThrLeuGlnLeuAspLysGluMetGlu 429
DB 2644 AGACACTCTTAGAACAGAGATTGAAGAGAGATGTTGGCTTTGCAGATGAACGCACA 2703
QY 430 GluLysAlaArgSerLeuSerLeuMetGluAlaAlaLeuAlaLysArgMetTyrAsn 449
DB 2704 GAACGAATACGT-----AGCCTGCTCGAGCGCCAGAGCCAGAGAAATTCAGACTTTGAC 2757
QY 450 GlnGlnIleLysArgProArgLeuSerProMetGluMetProProValThrSerSerSer 469
DB 2758 TCTGAAACGATGAGATAGTGTTTTGTATACATGCTCCTT-----TCTAAT 2802
QY 470 TyrSerProIleTyArgAspArgSerPheProSerGlnArgAspAspGlnAspGlu 489
DB 2803 CTCTCCCTCAGGCAATTCAGCCACAGCTACCCAGGA----- 2838
QY 490 IleSerAlaLeuValSerSerTyrLeuGlyProSerThrSerPheProHisArgSerArg 509
DB 2839 -----GCTTCTAGCTGTCTCACAAT----- 2859
QY 510 ArgSerProGluTyrMetValProLeuProHisGlyGlyLeuGlyArgSerValTyrAla 529
DB 2860 -----CCTACTGGGGTTCAGGA----- 2877
QY 530 TyrGluHisLeuAlaProAsnSerTyrSerProGlyHisGlyHisArgLeuHisArgGln 549
DB 2878 -----CCTCACTGGGGTTCATCCCATG----- 2898
QY 550 TyrSerProSerLeuValHisGlyGlnArgHisProLeuGlnTyrSerPro----- 566
DB 2899 -----GGTGGCACACCAAGCTTGGGGTTCATCCCATGCAAGGGCGGCCCAACCATGG 2952
QY 567 -----ProIleHisGlyGlnGlnGlnLeuProTyrGlyIleGlnArgValTyrArgHis 584
DB 2953 GGTCAACCCCTCAGGGCCAATGCAAGGGGTACCTCGAGGTAGCAGTATAGGAGTCCGCAAT 3012
QY 595 SerProSerGluGluArgTyrLeu-----GlyLeuSerAsnGlnArgSerProArg 601
DB 3013 AGCCCCCAGGCTCTGAGGCGGAGCAGCTTTCTGGGGGAGCGGACCGAACACAGGGCATGACAGA 3072
QY 602 SerAsnSer 604
DB 3073 AGCAGAGT 3081
RESULT 11
AAZ40489
ID AAZ40489 standard; DNA; 3824 BP.
XX

5080 GCCA 5083

RESULT 10
AAZ32435
ID AAZ32435 standard; cDNA; 3312 BP.
XX
AC AAZ32435;
XX
DT 27-JAN-2000 (first entry)
XX
DE Rat TAO1 kinase encoding cDNA.
XX
KW TAO1; MEK3; mitogen activated protein kinase; phosphorylation;
KW p38; protein kinase; cancer; inflammation; autoimmune disease;
KW degeneration; insulin-resistant diabetes; metabolic disorder;
KW neurodegeneration; MAP kinase; MAP/ERK kinase; ss.
XX
OS Rattus sp.
XX
FH Key Location/Qualifiers
FT CDS 121..3126
ET /*tag= a
FI /product= "TAO1 protein kinase"
XX
PN WO9953076-A1.
XX
PD 21-OCT-1999.
XX
PF 14-APR-1999; 99WO-US08165.
XX
PR 14-APR-1998; 98US-0060410.
XX
PS (TEXA) UNIV TEXAS SYSTEM.
XX
PI Cobb M, Hutchison M, Chen Z, Berman K;
XX
DR WPI, 1999-633831/54.
XX
DR P-PSDB; AAY49896.
XX
PT New polypeptides that phosphorylate kinase, used to screen for
PT modulators for treating e.g. cancer or inflammation -
XX
PS Claim 12; Fig 1; 95pp; English.
XX
CC The present sequence encodes rat TAO1 protein kinase, which is capable
CC of phosphorylating MEK3 (a MAP/ERK kinase). TAO kinases, and related
CC polypeptides, are used to screen for modulators of stress-responsive
CC mitogen activated protein (MAP) kinase pathways. These modulators are
CC potentially useful for treating or preventing: (1) inflammation,
CC autoimmune disease, cancer and degeneration (inhibitors of
CC phosphorylation); or (2) insulin-resistant diabetes, metabolic disorders
CC and neurodegeneration (enhancers of phosphorylation). TAO kinases are
CC also used to raise specific antibodies, useful therapeutically as
CC modulators and as immunoassay reagents for detecting TAO kinases.
CC TAO kinase polynucleotides can be used: (a) for recombinant expression
CC of TAO kinases; and (b) in the form of fragments, for detecting TAO
CC kinase polynucleotides in standard hybridization and amplification
CC tests. TAO kinases are highly specific for MEK3.
XX
SQ Sequence 3312 BP; 1106 A; 674 C; 780 G; 752 T; 0 other;
XX
Alignment Scores:
Pred. No.: 0.0257 Length: 3312
Score: 151.00 Matches: 150
Percent Similarity: 34.75% Conservative: 95
Best Local Similarity: 21.28% Mismatches: 237
Query Match: 4.86% Indels: 223
DB: 20 Gaps: 32
US-09-890-475-1 (1-609) x AAZ32435 (1-3312)
QY 3 AsnTyr-----ProProThrValAlaAlaGlnProThrThrAlaAsn 17
|||||
|||||

QY 455 ProArgLeuSerProMetGluMetProValThrSerSerSerTySerProIleTy 474
DB 1201 -----CCAAATGCCACCAAGAGCGGCGTATCACA----- 1233
QY 475 ArgAspArgSerPheProSerGlnArgAspAspGlnAspGluIleSerAlaLeuVal 494
DB 1234 -----AACGGTTACGTC 1245
QY 495 SerSerTyLeuGlyProSerThrSerPheProHisArgSerArgSerProGluTy 514
DB 1246 TCCCTTTCCACCGCCTCCACCAACATTCATCAGATCCCAAAATCACTCACCTCAATAC 1305
QY 515 MetValProLeuProHisGlyGlyLeuArgSerValTyAlaTyGluHisLeuAla 534
DB 1306 GGTGTACCA-----GCATACACCATCCCA 1332
QY 535 ProAnSerTy-----SerProGlyHisGlyHisArgLeuHisArgGlnTySer 551
DB 1333 CCTACCATCTACAGCAACAGAGCGCTCCATAC-----CAATACTCA 1374
QY 552 ProSerLeuValHisGly-----GlnArgHisProLeuGlnTySerProProIleHis 569
DB 1375 CCGAGGCGGTTCARGTCTTACAACTCTCCGTCTAGTTAT-----CCGACAGCATAT 1431
QY 570 GlyGlnGlnGlnLeuProTyxGlyle-----GlnArgValTyArgHisSerProSer 587
DB 1432 GGTACTACTGAGTCGCGTGGTCTCCACCCCTCCAGTTTACCATCTCCACCGCAC 1491
QY 588 GluGluArgTyLeu 592
DB 1492 CACCACCATAT 1506
RESULT 9
ABL17653
ID ABL17653 standard; DNA; 7785 BP.
XX ABL17653;
AC ABL17653;
XX 26-MAR-2002 (first entry)
DT Drosophila melanogaster genomic polynucleotide SEQ ID NO 4432.
DE Drosophila; developmental biology; cell signalling; insecticide;
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ds.
XX Drosophila melanogaster.
OS WO200171042-A2.
XX 27-SEP-2001.
PD 23-MAR-2001; 2001WO-US09231.
FF 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX (PEKE) PE CORP NY.
PA Venter JC, Adams M, Li PWD, Myers EW;
PI WPI; 2001-656860/75.
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX Claim 1; SEQ ID NO 4432; 21pp + Sequence Listing; English.
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of

CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 7785 BP; 2166 A; 2186 C; 2007 G; 1426 T; 0 other;
Alignment Scores:
Pred. No.: 0.0687 Length: 7785
Score: 152.00 Matches: 129
Percent Similarity: 37.78% Conservative: 106
Best Local Similarity: 20.74% Mismatches: 238
Query Match: 4.90% Indels: 149
DB: 23 Gaps: 31
US-09-890-475-1 (1-609) x ABL17653 (1-7785)
QY 5 ProProThrValAlaAlaGlnProThr-----ThrThrAlaAsnProLeuLeuGlnArgHis 23
DB 3487 CCTCGTGGTGGCGCAAGTCCAGCGCCGCAAGCAATCCTTTC----- 3534
QY 24 GlnSerGluGlnArgArgGluLeuProLysIleValGluThrGluSerThr----- 41
DB 3535 -----GACCAACCCCGCACCCGAGAGCCCGTCAATCCACCTATGCCACCCAGT 3588
QY 42 -----SerMetAspIleThrIleGlyGlnSerLys 51
DB 3589 CCTATCAGATGGAGGACACACCCCTGAGACCTGAGATCTCATTGGCAACATGGAA 3648
QY 52 GlnProGlnPheLeuLysSerIleAspGluLeuAlaAlaPheSerValAlaValGluThr 71
DB 3649 CAGAACCGCCCAAGCAGACACCGAGGTTCTACGTTCATCGAGACCATGCTGGACACA 3708
QY 72 PheLysArgGlnPheAspAspLeuGlnLysHisIleGluSerIleGluAsnAlaIleAsp 91
DB 3709 TTCAGGGAAAG-----CCGGAACCTGGAGTACGAAACCGAT 3744
QY 92 SerLysLeuGluSerAsnGlyValValLeuAlaAlaArgAsnAsnAsnPheHisGlnPro 111
DB 3745 CCGGAGATATATCGT-----ATCCTAACAAAGAGGCCCATAGCAATGAACCA 3795
QY 112 MetLeuSerProProArgAsnAsnValSerValGluThrThrValThr---ValSerGln 130
DB 3796 -----TCTGACCGCGCAGAGATAGTCAACCGATGATGTAAGAACCTTTGTATGGCAA 3849
QY 131 ProSerGlnGluIleValPro-----GluThrSerAsnLys 142
DB 3850 CACATTGAGGAAATTTGTGCCAACCTGTCACAGGAGGAGGCTCATGGAGAAA 3909
QY 143 ProGluGlyGlyArgMetCysGluLeuMetCysSerLysGlyLeuArgLysTyIleTy 162
DB 3910 CCGAAGAGGTAGCTAAATTGAG----- 3933
QY 163 AlaAsnIleSerAspGlnAlaLys-----LeuMetGluGluIleProSerAla 178
DB 3934 -----GCAGAACAGCCAAAGGAGCCTGAACCTCTACTCGAACCTCAGCCTGAGCCC 3984
QY 179 LeuLysLeuAlaLysGluProAlaLysPheValLeuAspCysIleGlyLysPheTyLeu 198
DB 3985 ACACCA---AGTGTGATCTCTGTCTACATCATGGAGGAGTCAAGCCCAATTTA 4041
QY 199 GlnGlyArgArgAlaPheThrLysGluSerProMetSerSer-----AlaArgGln 215
DB 4042 -----AGGAGAGCAGCATCTCGTGAAGATTTCAACCCAGTTTATCTATGCCACTGAA 4095
QY 216 ValSerLeuLeuIleLeuGluSerPheLeuLeuMetProAspArgGlyLysVal 235
DB 4096 ATAGCC-----AACTCAAGTTGGAATTTGATCGTGC----- 4128
QY 236 LysIleGluSerTrpIleLysAspGluAlaGluThrAlaAlaValAlaTrpArgLysArg 255

XX Arabidopsis thaliana; plant; gene; stress; transgenic; ds.
 XX Arabidopsis thaliana.
 OS Arabidopsis thaliana.
 XX WO200216655-A2.
 PN 28-FEB-2002.
 PD 24-AUG-2001; 2001WO-US26685.
 PP 24-AUG-2000; 2000US-227866P.
 PR 26-JAN-2001; 2001US-264647P.
 PR 22-JUN-2001; 2001US-300111P.
 XX (SCRI) SCRIPPS RES INST.
 PA (SYGN) SYNGENTA PARTICIPATIONS AG.
 XX Harper JF, Kreps J, Wang X, Zhu T;
 PI WPI; 2002-304127/34.
 XX
 XX Identifying a stress condition to which a plant cell has been exposed
 PT and producing plants with increased tolerance to these abiotic stresses
 PT
 XX
 XX Claim 144; SEQ ID NO 917; 577pp + Sequence Listing; English.
 PS
 XX The invention relates to identifying a stress condition to which a plant
 CC cell has been exposed, comprising:
 CC (a) contacting nucleic acid representative of expressed polynucleotides
 CC in the plant cell with an array or probes representative of the plant
 CC cell genome; and
 CC (b) detecting a profile of expressed polynucleotides in the plant cell
 CC characteristic of a stress response. The method is useful in the
 CC production of transgenic plants, cells and seeds and in producing plants
 CC with increased tolerance to abiotic stress. The present sequence is that
 CC of an Arabidopsis thaliana stress regulated gene (AB212196-AB212197) used
 CC in methods of the invention.
 CC Note: The sequence data for this patent is not represented in the printed
 CC specification but is based on sequence information supplied to Derwent by
 CC the European Patent Office.
 XX
 SQ Sequence 1524 BP; 401 A; 342 C; 391 G; 390 T; 0 other;

Alignment Scores:
 Pred. No.: 1.3e-09 Length: 1524
 Score: 234.50 Matches: 117
 Percent Similarity: 34.38% Conservative: 91
 Best Local Similarity: 19.34% Mismatches: 222
 Query Match: 7.55% Indels: 175
 Ds: 24 Gaps: 21

US-09-890-475-1 (1-609) x AB213112 (1-1524)

QY 49 GlnSerLysGlnProGlnPheLeuLysSerIleAspGluLeuAlaAlaPheSerValAla 68
 DB 34 AAGTCGTCTACCGCGAGTTCTTTGAG----- 60
 QY 69 ValGluThrPheLysArgGln-----PheAspAsp 78
 DB 61 -----TTTCAGAACGCGCTCTCTAATGACGAGCTGCAATCTTCTATGGAAGAG 111
 QY 79 LeuGlnLysHisIleGluSerIleGluAsnAlaIleAspSerLysLeuGluSerAsnGly 98
 DB 112 CTCTCTGAGCAATTCACCTTCATGCGACAAACCTGATGAGAAATCTGAGCTCTGAGA 171
 QY 99 ValValLeuAlaAlaArgAsnAsnPheHisGlnProMetLeuSerProProArgAsn 118
 DB 172 CAGATGATTGAACCCCTAGATAATCAACGACGAGCTCGATTGATTGATTCTCAAGCATCGT 231
 QY 119 AsnValSerValGluThrThrValThrValSer----- 129

DB 232 GAGGTCACTATAGACCAACAGCGTCGAGATCGCGAGGGGAAAGTTGAGAACCGGTTAGA 291
 QY 130 -----GlnProSerGlnGluLeuValProGluThrSerAsnLysPro 143
 DB 292 GCGCGCTTGGATTGCTGAGAAAGCTAGAGACTGTGTGTATGATACACCGGAGAGTC 351
 QY 144 GluGlyGly-----ArgMetCysGluLeuMetCysSerLys 155
 DB 352 GACGACGGGACGGGCTTTATCGCGCGCTGAAGTCGCTTTGTTTGAAGATGAGCGCAAGA 411
 QY 156 GlyLeuArgLysTyrIlePheAlaAsnIleSerAspGlnAlaLysLeuMetGluGluLe 175
 DB 412 GGATTTTGGGGTTTGTGATTCCGAGGAAGGAATTCGGAATCTCCGCTCACAGATT 471
 QY 176 ProSerAlaLeuLysLeuAlaLysGluProAlaLysPheValLeuAspCysIleGlyLys 195
 DB 472 CCTGCGCGTTGGTAGATTGTGTGATCCGCCGAAGTTAGTACTTGAAGCTGTATCTGAG 531
 QY 196 PheTyrLeuGlnGlyArgAlaPheThrLysGluSerProMetSerSerAlaArgGln 215
 DB 532 GTTTTTCGCGTGGATAAAAGAGGT---GGTGAGAGAAAGTCAGCAATGATTTTGGGATGG 588
 QY 216 ValSerLeuLeuLeuLeuSerPhe-----LeuLeuMetProAspArgGlyLys 232
 DB 589 GCTTGTGTGTGATCTGAGAGATTGATCCAGTTATGGTTGATCCAGTGTGGGAAG 648
 QY 233 GlyLysValLysIleGluSerTyrLysAspGluAlaGluThrAlaAlaValAlaTyr 252
 DB 649 TCTCGCTCTTGTAACTCTAGCTTAAGGAGAAAGCTAAAGAGATTGCTGAGACGTGG 708
 QY 253 ArgLysArgLeuMetThrGluGlyLeuAlaAlaGluLysMetAspAlaArgGly 272
 DB 709 AAGCTAGCTTGAAGAGAGGAGGATTGAGATGTGAACACACCTGATGTTCTATACG 768
 QY 273 LeuLeuLeuLeuValAlaCysPheGlyValProSerAsnPheArgSerThrAspLeu 292
 DB 769 TTTCTGCAATCTTGTGACTTTTGGATTGTGAAGAAG-----GATGATCTA 816
 QY 293 AspLeuLeuArgMetSerGlySerAsnGluLeuAlaGlyAlaLeuLysArgSerGlnPhe 312
 DB 817 GCTCTTTATAGG-----AAACTGTGTGTGTGCTAGCTTGGCGTGAACAGATG 864
 QY 313 -----LeuValProMetValSerGlyIleValGluSerSer 324
 DB 865 CCAAGCTTCTTTCAGTTGGTTGGCGCAAAATGCTGTGATGATTGAAGAGTTG 924
 QY 325 IleLysArgGlyMetHisIleGluAlaLeuGluMetValTyrThrPheGlyMetGluAsp 344
 DB 925 ATAATCAGGGGACACACAGCTTGTGCGGTTCTATTTCACTTTTGAAGTTGGTCTTTACAC 984
 QY 345 LysPheSerAlaAlaLeuValLeuThrSerPheLeuLysMetSerLysGluSerPheGlu 364
 DB 985 TTGTTCCCTCCTCTTCTTGTCTCAAGCTTATCTGAGGGATGCAAGAAAGCTACAGCT 1044
 QY 365 -----ArgAlaLysArgLysAlaGlnSerProLeu 374
 DB 1045 TTAATCACTGATGATCCAAACAAATCTGCGCGATCTGCGGGATCTGTTCTCTCAGCTC 1104
 QY 375 AlaPheLysGluAlaAlaThrLysGlnLeuAlaValLeuSerSerValMetGlnCysMet 394
 DB 1105 GCTTAC-----CAATGTATG 1119
 QY 395 GluThrHisLysLeuAspProAlaLysGluLeuProGlyTyrGlnIleLysGluGlnIle 414
 DB 1120 TGTCTAGAGAAAGCAAAACCGAGAGAGAAAAACCAAGCC-----GTG 1161
 QY 415 ValSerLeuGluLysAspThrLeuGlnLeuAspLysGluMetGluLysAlaArgSer 434
 DB 1162 ATCCGCCCAACAGAGAACCC----- 1182
 QY 435 LeuSerLeuMetGluGluAlaAlaLeuAlaLysArgMetTyrAsnGlnGlnIleLysArg 454
 DB 1183 -----CGAGCAAGCTACAGCGGT----- 1200

Pred. No.:	6,51e-14	Length:	1865
Score:	288.50	Matches:	132
Percent Similarity:	36.95%	Conservative:	103
Best Local Similarity:	20.75%	Mismatches:	250
Query Match:	9.29%	Indels:	151
DB:	21	Gaps:	21

JS-09-890-475-1 (1-609) x RAC38630 (1-1865)

2y	37	GluThrGluSerThrSerMetAspIleThrIleGlyGlnSerLysGlnProGlnPheLeu	56
Db	68	GAGATGGAGTCTTCTCCGAGTCCAGGAGAGTTGATTAAGTCGTCAGCGAGTTTCTTT	127
2y	57	LysSerIleAspGluLeuAlaPheSerValAlaValGluThrPheLysArgGln---	75
Db	128	GAG-----TTTCAGAGCAGCGC	145
2y	76	-----PheAspAspLeuGlnLysIleGluSerIle	86
Db	146	TCTCTAATGACGAGTGCATCTTCTATGGAAGAGCTCTGAGCATTTCACTTCCATG	205
2y	87	GluAsnAlaIleAspSerLysLeuGluSerAsnGlyValValLeuAlaAlaArgAsnAsn	106
Db	206	GAGCAAAACCTGATGAGAATCTGAAGCTCTGAGACAGATGATGAACCCCTAGATAT	265
2y	107	AsnPheHisGlnProMetLeuSerProProArgAsnAsnValSerValGluThrThrVal	126
Db	266	CAAACGACGAGTCGATTTAGTTACTCAAGCATCGTGAGGTGACTATAGACACACGCTC	325
2y	127	ThrValSer-----GlnPro	131
Db	326	GAGATCGCGGAGGAAAGTTGAGAACCGCTTAGACGCGGTTGGATTTCGTGGAGAA	385
2y	132	SerGlnGluIleValProGluThrSerAsnLysProGluGlyGly-----	146
Db	386	GCTAGAGACTGTGTGATGAGATACCGGAGAACTCGACGCGCGCGGCTTTTATCG	445
2y	147	-----ArgMetCysGluLeuMetCysSerLysGlyLeuArgLysThrIleThrAla	163
Db	446	GCGCTGAATGCTGTTGTTGAAGATGACCAAGAGGATTTTGGGGGTTTGTGATGCG	505
2y	164	AsnIleSerAspGlnAlaLysLeuMetGluIleProSerAlaLeuLysLeuAlaLys	183
Db	506	AGGAAGAGAAATGGAGAACTCCGTCACAGATTCCTGTGGGCTTGGTAGATTGTG	565
2y	184	GluProAlaLysPheValLeuAspCysIleGlyLysPheThrLeuGlnGlyArgAla	203
Db	566	GATCCGCGCAAGTTAGTACTTGAAGCTGATCTGAGGTTTTTCCGGTGGATAAAAGAGGT	625
2y	204	PheThrLysGluSerProMetSerSerAlaArgGlnValSerLeuLeuGluSer	223
Db	626	---GGTGAGAGAAAGTAGCAATGTTTGGATGGGCTGTGTGTGTGATCTGGAGAGT	682
2y	224	Phe-----LeuLeuMetProAspArgGlyLysGlyLysValIleGluSerTrp	240
Db	683	TTGATTCAGTTATGTTGATTCAGTATGAGGAGTCTCGGCTACTTGTAACTCTAGC	742
2y	241	IleLysAspGluAlaGluThrAlaAlaValAlaTrpArgLysArgLeuMetThrGluGly	260
Db	743	GTTAAGGAGAAAGCTTAAAGAGATTCCTGAGCGTGGAGGCTAGCTTGGAGAGAGAGGA	802
2y	261	GlyLeuAlaAlaGluLysMetAspAlaArgGlyLeuLeuLeuValAlaCysPhe	280
Db	803	GGGATTGAGAATGTGAACACCTGATGTTTCATAGTTTCTGCAACATCTTGTGACITTT	862
2y	281	GlyValProSerAsnPheArgSerThrAspLeuLeuAspLeuIleArgMetSerGlySer	300
Db	863	GGGATTGTGAAGAAG-----GATGATCTAGTCTCTTTATAGG-----	898
2y	301	AsnGluIleAlaGlyAlaLeuLysArgSerGlnPhe-----	312
Db	899	AAACTTGTGGTTGCTTACGTTGGGCTAAACAGATGCCAAGCTGCTGTTCAGTTGGT	958

Qy	313	LeuValProMetValSerGlyIleValGluSerSerIleLysArgGlyMetHisIleGlu	332
Db	959	TTGGCGACCAATGCTGATATGATTTGAAGTTGATATCAGGGGCAACAGCTTGAT	1018
Qy	333	AlaLeuGluMetValThrThrPheGlyMetGluAspLysPheSerAlaAlaLeuValLeu	352
Db	1019	CGGTTTCATTTCACTTTTGAAGTTGCTTTGACACTTGTTCCTCTCTGCTTGTCTC	1078
Qy	353	ThrSerPheLeuLysMetSerLys-----GluSerPheGluArg	365
Db	1079	AAAGCTTATCTGAGGATGCAAAAGAGCTACAGCTTTTATCACTGATGATCCACAAAT	1138
Qy	366	AlaLysArgLysAlaGlnSerProLeuAlaPheLysGluAlaThrLysGlnLeuAla	385
Db	1139	TCTGGCGATCTGGCATCTT-----GTGGCGGCAAGAGCAATCA	1180
Qy	386	ValLeuSerSerValMetGlnCysMetGluThrHisLysLeuAspProAlaLysGluLeu	405
Db	1181	GCATTTAGACGATCTTAAATGATAGAGATCAAACTCGAG-----GAAGAATTC	1234
Qy	406	ProGlyTrpGlnIleLysGluGlnIleValSerLeuGluLysAspThrLeuGlnLeuAsp	425
Db	1235	CCACCCGAGATCTCAAGAAGAGGTTGGATCAGTAGAGAGCAACCAACCCGAGAAGA	1294
Qy	426	LysGluMetGluLysAlaArgSerLeuSerLeuMetGluAlaAlaLeuAlaLys	445
Db	1295	AAACCCGCGGTGATTCGCGCAAC-----	1318
Qy	446	ArgMetTyrAsnGlnGlnIleLysArgProArgLeuSer---ProMetGluMetProPro	464
Db	1319	-----NAGAGAACCCGAGCAAGCTACAGCGGTCCAATGCCACCA	1357
Qy	465	ValThrSerSerSerTyrSerProIleTyrArgAspArgSerPheProSerGlnArgAsp	484
Db	1358	GCCAAAGCAGCGGCTATCACA-----	1378
Qy	485	AspAspGlnAspGluIleSerAlaLeuValSerSerTyrLeuGlyProSerThrSerPhe	504
Db	1379	-----AACGTTAGCTCTCTCTTCCACCGCTCCACCAATTC	1420
Qy	505	ProHisArgSerArgSerProGluTyrMetValProLeuProHisGlyGlyLeuGly	524
Db	1421	ATCAGATCCCAAGTCACTCACCTCAATACGTTGTACCA-----	1459
Qy	525	ArgSerValTyrAlaTyrGluHisLeuAlaProAsnSerTyr-----SerProGly	541
Db	1460	-----GCATACACCATCCCTCCATCTTACAGCAACGAGGAGCCCTCCA	1507
Qy	542	HisGlyHisArgLeuHisArgGlnTyrSerProSerIleValHisGly-----GlnArg	559
Db	1508	TAC-----CAATCTACCCGAGCGGTTTCATGTTCTTACCAAAACC	1549
Qy	560	HisProLeuGlnTyrSerProIleHisGlyGlnGlnLeuProTyrGlyIle---	578
Db	1550	TCTCGGTTCACTTAT---CCGACAGCATATGGTACCTACTGCGAGTCCGGTGGCTGCCA	1606
Qy	579	-----GlnArgValTyrArgHisSerProSerGluGluArgTyrLeu	592
Db	1607	CCCCCTCCAGTTTACCCTCTCCCGCACCCACCCATATTCAGCATGTTTACTTAC	1666
Qy	593	GlyLeuSerAsnGlnArgSerProArgSerAsnSerSerLeuAspPro	608
Db	1667	TGATTAGACAATGACAAAGGAAAAAGAGATGATGTAAGTAACCAACCA	1714

RESULT 8
AB213112
ID AB213112 standard; DNA; 1524 BP.
XX AC AB213112;
XX DT 21-JAN-2003 (first entry)
XX DE Arabidopsis thaliana stress regulated gene SEQ ID NO 917.

PR	03-JUN-1999;	99US-0137528.	PR	05-AUG-1999;	99US-0147260.
PR	04-JUN-1999;	99US-0137502.	PR	06-AUG-1999;	99US-0147303.
PR	07-JUN-1999;	99US-0137724.	PR	06-AUG-1999;	99US-0147416.
PR	08-JUN-1999;	99US-0138094.	PR	09-AUG-1999;	99US-0147493.
PR	10-JUN-1999;	99US-0138540.	PR	09-AUG-1999;	99US-0147935.
PR	10-JUN-1999;	99US-0138847.	PR	10-AUG-1999;	99US-0148171.
PR	10-JUN-1999;	99US-0139119.	PR	11-AUG-1999;	99US-0148319.
PR	14-JUN-1999;	99US-0139119.	PR	12-AUG-1999;	99US-0148341.
PR	16-JUN-1999;	99US-0139452.	PR	13-AUG-1999;	99US-0148565.
PR	16-JUN-1999;	99US-0139453.	PR	13-AUG-1999;	99US-0148684.
PR	17-JUN-1999;	99US-0139492.	PR	16-AUG-1999;	99US-0149368.
PR	18-JUN-1999;	99US-0139454.	PR	17-AUG-1999;	99US-0149175.
PR	18-JUN-1999;	99US-0139456.	PR	18-AUG-1999;	99US-0149426.
PR	18-JUN-1999;	99US-0139456.	PR	20-AUG-1999;	99US-0149426.
PR	18-JUN-1999;	99US-0139457.	PR	20-AUG-1999;	99US-0149723.
PR	18-JUN-1999;	99US-0139458.	PR	20-AUG-1999;	99US-0149723.
PR	18-JUN-1999;	99US-0139459.	PR	23-AUG-1999;	99US-0149929.
PR	18-JUN-1999;	99US-0139460.	PR	23-AUG-1999;	99US-0149902.
PR	18-JUN-1999;	99US-0139461.	PR	23-AUG-1999;	99US-0149930.
PR	18-JUN-1999;	99US-0139461.	PR	25-AUG-1999;	99US-0150566.
PR	18-JUN-1999;	99US-0139462.	PR	26-AUG-1999;	99US-0150884.
PR	18-JUN-1999;	99US-0139463.	PR	27-AUG-1999;	99US-0151065.
PR	18-JUN-1999;	99US-0139463.	PR	27-AUG-1999;	99US-0151066.
PR	18-JUN-1999;	99US-0139750.	PR	27-AUG-1999;	99US-0151066.
PR	18-JUN-1999;	99US-0139763.	PR	27-AUG-1999;	99US-0151080.
PR	21-JUN-1999;	99US-0139817.	PR	27-AUG-1999;	99US-0151303.
PR	22-JUN-1999;	99US-0139899.	PR	30-AUG-1999;	99US-0151303.
PR	23-JUN-1999;	99US-0140353.	PR	31-AUG-1999;	99US-0151438.
PR	23-JUN-1999;	99US-0140353.	PR	01-SEP-1999;	99US-0151930.
PR	23-JUN-1999;	99US-0140354.	PR	07-SEP-1999;	99US-0152363.
PR	24-JUN-1999;	99US-0140635.	PR	10-SEP-1999;	99US-0153070.
PR	28-JUN-1999;	99US-0140823.	PR	13-SEP-1999;	99US-0153758.
PR	29-JUN-1999;	99US-0140991.	PR	16-SEP-1999;	99US-0154018.
PR	30-JUN-1999;	99US-0141287.	PR	16-SEP-1999;	99US-0154039.
PR	01-JUL-1999;	99US-0141842.	PR	20-SEP-1999;	99US-0154779.
PR	01-JUL-1999;	99US-0142154.	PR	22-SEP-1999;	99US-0155139.
PR	02-JUL-1999;	99US-0142055.	PR	23-SEP-1999;	99US-0155486.
PR	06-JUL-1999;	99US-0142300.	PR	24-SEP-1999;	99US-0155659.
PR	08-JUL-1999;	99US-0142803.	PR	28-SEP-1999;	99US-0156458.
PR	09-JUL-1999;	99US-0142920.	PR	29-SEP-1999;	99US-0156596.
PR	12-JUL-1999;	99US-0142977.	PR	04-OCT-1999;	99US-0157117.
PR	13-JUL-1999;	99US-0143542.	PR	05-OCT-1999;	99US-0157753.
PR	14-JUL-1999;	99US-0143624.	PR	06-OCT-1999;	99US-0157865.
PR	15-JUL-1999;	99US-0144005.	PR	07-OCT-1999;	99US-0158029.
PR	16-JUL-1999;	99US-0144085.	PR	08-OCT-1999;	99US-0158232.
PR	16-JUL-1999;	99US-0144086.	PR	12-OCT-1999;	99US-0158369.
PR	19-JUL-1999;	99US-0144325.	PR	13-OCT-1999;	99US-0159293.
PR	19-JUL-1999;	99US-0144331.	PR	13-OCT-1999;	99US-0159294.
PR	19-JUL-1999;	99US-0144332.	PR	13-OCT-1999;	99US-0159295.
PR	19-JUL-1999;	99US-0144333.	PR	14-OCT-1999;	99US-0159329.
PR	19-JUL-1999;	99US-0144334.	PR	14-OCT-1999;	99US-0159330.
PR	19-JUL-1999;	99US-0144335.	PR	14-OCT-1999;	99US-0159331.
PR	20-JUL-1999;	99US-0144352.	PR	14-OCT-1999;	99US-0159331.
PR	20-JUL-1999;	99US-0144632.	PR	14-OCT-1999;	99US-0159637.
PR	20-JUL-1999;	99US-0144884.	PR	14-OCT-1999;	99US-0159638.
PR	21-JUL-1999;	99US-0144814.	PR	18-OCT-1999;	99US-0159584.
PR	21-JUL-1999;	99US-0145086.	PR	21-OCT-1999;	99US-0160741.
PR	21-JUL-1999;	99US-0145088.	PR	21-OCT-1999;	99US-0160767.
PR	21-JUL-1999;	99US-0145088.	PR	21-OCT-1999;	99US-0160768.
PR	22-JUL-1999;	99US-0145085.	PR	21-OCT-1999;	99US-0160770.
PR	22-JUL-1999;	99US-0145087.	PR	21-OCT-1999;	99US-0160814.
PR	22-JUL-1999;	99US-0145089.	PR	21-OCT-1999;	99US-0160815.
PR	22-JUL-1999;	99US-0145192.	PR	21-OCT-1999;	99US-0160815.
PR	23-JUL-1999;	99US-0145145.	PR	22-OCT-1999;	99US-0160980.
PR	23-JUL-1999;	99US-0145218.	PR	22-OCT-1999;	99US-0160981.
PR	23-JUL-1999;	99US-0145224.	PR	22-OCT-1999;	99US-0160981.
PR	23-JUL-1999;	99US-0145224.	PR	25-OCT-1999;	99US-0161404.
PR	26-JUL-1999;	99US-0145276.	PR	25-OCT-1999;	99US-0161405.
PR	27-JUL-1999;	99US-0145913.	PR	25-OCT-1999;	99US-0161406.
PR	27-JUL-1999;	99US-0145918.	PR	25-OCT-1999;	99US-0161406.
PR	27-JUL-1999;	99US-0145919.	PR	26-OCT-1999;	99US-0161359.
PR	28-JUL-1999;	99US-0145951.	PR	26-OCT-1999;	99US-0161360.
PR	02-AUG-1999;	99US-0146388.	PR	26-OCT-1999;	99US-0161361.
PR	02-AUG-1999;	99US-0146388.	PR	28-OCT-1999;	99US-0161920.
PR	02-AUG-1999;	99US-0146389.	PR	28-OCT-1999;	99US-0161992.
PR	03-AUG-1999;	99US-0147038.	PR	28-OCT-1999;	99US-0161993.
PR	04-AUG-1999;	99US-0147204.	PR	29-OCT-1999;	99US-0162142.
PR	04-AUG-1999;	99US-0147302.			
PR	05-AUG-1999;	99US-0147192.			

Alignment Scores:

QY 162 TyrAlaAsnIleSerAspGlnAlaLysLeuMetGluGluIleProSerAlaLeuLysLeu 181
DB 545 ACGCGAGGAGAGAGGATTGGAGATCTCCGGTCAAGGATTCGGCGGCTTGGTTGAT 604
QY 182 AlAlaLysGluProAlaLysPheValLeuAspCysIleGlyLysPheTyrLeuGlnGlyArg 201
DB 605 TGTGTAGATCCGGGATGTTAGTGTCTGAAGCGATATCTGAGGTTTTTCGGTGGATACA 664
QY 202 ArgAlaPheThrLysGluSerProMetSerSerAlaArgGlnValSerLeuLeuLeu 221
DB 665 AGAGAGATAAAGTTAGCAATATTATGATGGGCTTGTGTGTGATCTTGAGAGTTTA 724
QY 222 GluSerPheLeuLeuMetProAspArgLysGlyLysValLysIleGluSerTrpIle 241
DB 725 ACACCTGTAATTGTTGATCCATGATGGGAATCGAGGTTGCTTGTACTCCGAGTGT 784
QY 242 LysAspGluAlaGluThrAlaAlaValAlaIleTrrpGlyLysArgLeuMetThrGluGlyGly 261
DB 785 AAGGAGAAAGCTTAAGGAGATTGCTGAGACGTGGGAAGAGATTGGAAGAGAGAGGAGG 844
QY 262 LeuAlaAlaAlaGluLysMetAspAlaAlaArgGlyLeuLeuLeuValAlaCysPheGly 281
DB 845 ATAGAGATGTGAACACTCTGATGTTTCATCGTTTTTGCAGCATCTGTATATTGGA 904
QY 282 ValProSerAsnPheArgSerThrAspLeuLeuAspLeuLeuLeuArgMetSerGlySerAsn 301
DB 905 ATT-----GTTAAAGTCAAGATCTTCTGTTGTATAGAAAGCTT----- 943
QY 302 GluIleAlaGlyAlaLeuLysArgSerGlnPhe-----Leu 313
DB 944 ---GTTGTGGATCTGCGTGGGTAACAGATGCTTAAGCTTGGCGTTTCGTGGTTG 1000
QY 314 ValProMetValSerGlyLeuValGluSerSerIleLysArgGlyMetHisIleGluAla 333
DB 1001 GGTGATCAATCCCTGCATGATGATTGAAGATTAATATCAGCGGGACACAGCTTGATGCT 1060
QY 334 LeuGluMetValTyrPheGlyMetGluAspLysPheSerAlaAlaLeuValLeuThr 353
DB 1061 GTTCATTTTACTATTGAAGTTGGGCTGTGGATAAGTTCCCTCTGTTCTCTTGTCTCAAA 1120
QY 354 SerPheLeuLysMetSerLysGluSerPheGluArg---AlaLysArgLysAlaGlnSer 372
DB 1121 GCTTATCAAGGACGCAAGAAAGTCAGCAGCTTCTATCATGGAGGACTCTAGCAATACT 1180
QY 373 ProLeuAlaPheLysGluAlaAlaThrLysGlnLeuAlaValLeuSerSerValMetGln 392
DB 1181 GGCAGGCTACGATCTCTGTCGCGCAAGGAGCAATCAGCATTAAAGCGGTTTAAAA 1240
QY 393 CysMetGluThrHisLysLeuAspProAlaLysGluLeuProGlyTrrpGlnIleLysGlu 412
DB 1241 TGCATAGAAGATCAAACTCGAG-----GAAGAGTTCCCTCCAGAGAACCCTCAAGAA 1294
QY 413 GlnIleValSerLeuGluLysAspThrLeuGlnLeuAspLysGluMetGluGluLysAla 432
DB 1295 CGGTTG-----GATCAGCTAGAGAA-GACCAAAACCGAGAGAGAAAGCC 1338
QY 433 ArgSerLeuSerLeuMetGluGluAlaLeuAlaLysArgMetTyrAsnGlnIle 452
DB 1339 AGCAGCTGTCTTCCCAACAAAGAACCCG-----AGCCAGCTA 1377
QY 453 LysArg-ProArgLeuSerProMetGluMetProValThrSerSerSerTyrSerPr 472
DB 1378 CAACGGTCCA-----ATGCCACCAAGGCGCGGCGGTATCACAAA 1419
QY 472 IleTyrArgAspArgSerPheProSerGlnArgAspAspGlnAspGluIleSerAl 492
DB 1420 CGCATACGTCCTT---TCCTTCCCTTTCATCAGATCACCTCTCCTCTCTCTCAATACGC 1476
QY 492 aLeuValSerSerTyrLeuGlyProSerThrSerPheProHisArgSerArgSerPr 512
DB 1477 TTCACGAGGAGCTTACCATCCCACTTACACAGTC-----TACAGCAACAGAGCC 1530
QY 512 oGluTyr-----MetValProLeuProHisGlyLysLeuGlyArgSerVa 527

DB 1531 ACCTTATCCATCTCTCCGAGATCATCCAGGCTCATACCAAGGCTCTCTATCGTTA 1590
QY 527 lTyrAlaTyrGluHisLeu-----AlaProAsnSerTyrSe 539
DB 1591 CCCAGCGTACAACGGTTATTGTAAATGCTCCAGTCTCTGCTCCAGCTCTCCAGTTTACCA 1650
QY 539 rProGlyHisGlyHisArgLeuHisArgGlnTyrSerProSerLeuValHisGlyGlnAr 559
DB 1651 CCCA---CACCATCATCAA---CACCACCAATTC-----CACCATCAGCA 1689
QY 559 gHis 560
DB 1690 ACAT 1693
RESULT 7
AAC38630
ID AAC38630 standard; DNA; 1865 BP.
XX AAC38630;
XX 17-OCT-2000 (first entry)
XX Arabidopsis thaliana DNA fragment SEQ ID NO: 21672.
XX Hybridisation assay; genetic mapping; gene expression control;
XX protein identification; signal transduction pathway;
XX metabolic pathway; promoter; termination sequence; ss.
XX Arabidopsis thaliana.
XX EP1033405-A2.
XX 06-SEP-2000.
XX 25-FEB-2000; 2000EP-0301439.
XX 25-FEB-1999; 99US-0121825.
XX 05-MAR-1999; 99US-0123180.
XX 09-MAR-1999; 99US-0123548.
XX 23-MAR-1999; 99US-0125788.
XX 25-MAR-1999; 99US-0126264.
XX 29-MAR-1999; 99US-0126785.
XX 01-APR-1999; 99US-0127462.
XX 06-APR-1999; 99US-0128234.
XX 08-APR-1999; 99US-0128714.
XX 16-APR-1999; 99US-0129845.
XX 19-APR-1999; 99US-0130077.
XX 21-APR-1999; 99US-0130449.
XX 23-APR-1999; 99US-0130510.
XX 28-APR-1999; 99US-0130891.
XX 30-APR-1999; 99US-0131449.
XX 30-APR-1999; 99US-0132048.
XX 30-APR-1999; 99US-0132407.
XX 04-MAY-1999; 99US-0132484.
XX 05-MAY-1999; 99US-0132485.
XX 06-MAY-1999; 99US-0132486.
XX 07-MAY-1999; 99US-0132487.
XX 07-MAY-1999; 99US-0132863.
XX 11-MAY-1999; 99US-0134256.
XX 14-MAY-1999; 99US-0134218.
XX 14-MAY-1999; 99US-0134219.
XX 14-MAY-1999; 99US-0134221.
XX 14-MAY-1999; 99US-0134370.
XX 18-MAY-1999; 99US-0134768.
XX 19-MAY-1999; 99US-0134941.
XX 20-MAY-1999; 99US-0135124.
XX 21-MAY-1999; 99US-0135353.
XX 24-MAY-1999; 99US-0135629.
XX 25-MAY-1999; 99US-0136021.
XX 27-MAY-1999; 99US-0136392.
XX 28-MAY-1999; 99US-0136782.
XX 01-JUN-1999; 99US-0137222.

```
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 17-JUL-1999; 99US-0144086.
PR 18-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 28-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 06-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151436.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.

PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Alignment Scores:
Pred. No.: 1,55e-14
Score: 296.50
Percent Similarity: 41.10%
Best Local Similarity: 23.13%
Query Match: 9.55%
DB: 21

US-09-890-475-1 (1-609) x AAC35353 (1-1970)
Qy 59 IleAspGluLeuAlaAlaPheSerValAlaValGluThrPheLysArgGln----- 75
Db 131 TTAAGTGGTTAGCT-----CAGCCGAGCTTCGAGGAGCTTTCAGAAACAGACGTCGTTG 184
Qy 76 -----PheAspAspLeuGlnLysHisIleGluSerIleGluAsn 88
Db 185 ATGACTAGCTGTACTCTTCTTATGGCAAGAGCTCTCCGATCACTTCCTCTCTAGAGCAG 244
Qy 89 AlaIleAspSerLysLeuGluSerAsnGlyValValLeuAlaAlaArgAsnAsnPhe 108
Db 245 AATCTGATGAAGAATACTGAGCGTTGAAGCAGATGATTGAACCTAGATAACCCAGACT 304
Qy 109 HisGlnProMetLeuSerProArgAsnAsnValSerValGluThrThrValThrVal 128
Db 305 CAGACCTCGCTCGAGTCTCTGAAGCGTAGAGAGGTAAACAATCGACCATAGCGTTGAGATC 364
Qy 129 -----SerGlnProSerGln 133
Db 365 GTAGCTGGGAAAGTAGGTGAACGAGCTAGAGTGTCTTGAATCGTTAGAGAAAGCTAGA 424
Qy 134 GluIleValProGluThrSerAsnLysProGluGlyGly----- 146
Db 425 GATGGTTGTGTGATGGTTCTTAATGATGATTCGGTGATGTTCACGATGACGATGAGAGGTCTT 484
Qy 147 -----ArgMetCysGluLeuMetCysSerLysGlyLeuArgLysTyrIle 161
Db 485 CTCCTCTGCTTTGAAATCTCTTTGTCTTAAATGATGATGCGGAGAGGTTTTTGGAACTTTGTG 544
```


PR	06-AUG-1999;	99US-0147303.	Score:	296.50	Matches:	130
PR	06-AUG-1999;	99US-0147416.	Percent Similarity:	41.10%	Conservative:	101
PR	09-AUG-1999;	99US-0147493.	Best Local Similarity:	23.13%	Mismatches:	231
PR	09-AUG-1999;	99US-0147935.	Query Match:	9.55%	Indels:	101
PR	10-AUG-1999;	99US-0148171.	DB:	21	Gaps:	19
PR	11-AUG-1999;	99US-0148319.	US-09-890-475-1 (1-609) x AAC49876 (1-1964)			
PR	12-AUG-1999;	99US-0148341.	QY	59	IleAspGluLeuAlaAlaPheSerValAlaValGluThrPheLysArgGln-----	75
PR	13-AUG-1999;	99US-0148565.	DB	128	TTAACTGAGTTAGCT-----CAGCCGAGTTTCGAGGAGTTTCAGAAACACACAGCTCGTTG	181
PR	16-AUG-1999;	99US-0149368.	QY	76	-----PheAspSerLeuGlnLysHisIleGluSerIleGluAsn	88
PR	17-AUG-1999;	99US-0149175.	DB	182	ATGACTAGCTGACTCTTCTATGCGAAGAGCTCCGATCCTCCTCCTCTCTAGAGCAG	241
PR	18-AUG-1999;	99US-0149426.	QY	89	AlaIleAspSerLysLeuGluSerAsnGlyValValLeuAlaAlaArgAsnAsnAsnPhe	108
PR	20-AUG-1999;	99US-0149723.	DB	242	ATCTGATGAAGAAATCTGAAGCGTTGAAGCAGATGATTGAACCTCTAGATACACGACT	301
PR	20-AUG-1999;	99US-0149722.	QY	109	HisGlnProMetLeuSerProProArgAsnValSerValGluThrValThrVal	128
PR	23-AUG-1999;	99US-0149902.	DB	302	CAGACCTCGCTCGAGTCTCTGAAGCGTAGAGAGGTAAACAATCGACCATAGCGTTGAGATC	361
PR	25-AUG-1999;	99US-0149930.	QY	129	-----SerGlnProSerGln	133
PR	26-AUG-1999;	99US-0150566.	DB	362	GTAGCTGGGAAGTAGGTGAACGACTAGAGCTGCTCTTGAATCGTTAGAGAAAGTAGA	421
PR	27-AUG-1999;	99US-0150884.	QY	134	GluIleValProGluThrSerAsnLysProGluGlyGly-----	146
PR	27-AUG-1999;	99US-0151065.	DB	422	GATGTTGTGTGATGCTTCTAATGATGATTCGGTGATGTTGACGATGAAGAAGGTCTT	481
PR	27-AUG-1999;	99US-0151080.	QY	147	-----ArgMetCysGluLeuMetCysSerLysGlyLeuArgLysTrpIle	161
PR	28-AUG-1999;	99US-0151403.	DB	482	CTTTCTGCTTTGAAATCTCTTTGCTTAAATGGAATCGAGAGGTTTGGAACTTTGTG	541
PR	28-AUG-1999;	99US-0151438.	QY	162	TyrAlaAsnIleSerAspGlnAlaLysLeuMetGluIleProSerAlaLeuLysLeu	181
PR	28-SEP-1999;	99US-0151513.	DB	542	ACGCGAGGAGAGAGGAGTGGAGATCTCCGGTCAAAGATTCGCGCGCGTGTGTTGAT	601
PR	28-SEP-1999;	99US-0151537.	QY	182	AlaLysGluProAlaLysPheValLeuAspCysIleGlyLysPheLysLeuGlnGlyArg	201
PR	28-SEP-1999;	99US-0151540.	DB	602	TGTGTAGATCCGCGGATGTTAGTGTCTGAAGCGATATCTAGGTTTTCGCGTGGATACA	661
PR	28-SEP-1999;	99US-0151543.	QY	202	ArgAlaPheThrLysGluSerProMetSerSerAlaArgGlnValSerLeuLeuLeu	221
PR	28-SEP-1999;	99US-0151547.	DB	662	AGAGGAGATAAAGTTAGCAATGATTATGCGTGTGTTGTTGATCTTCTGAGAGTTTA	721
PR	28-SEP-1999;	99US-0151548.	QY	222	GluSerPheLeuLeuMetProAspArgGlyLysGlyLysValLysIleGluSerTrpIle	241
PR	28-SEP-1999;	99US-0151549.	DB	722	ACACCTGTAATTCTTGTATCCAGTATGGGAATCGAGTTGCTTGTACTCCGAGTGT	781
PR	28-SEP-1999;	99US-0151550.	QY	242	LysAspGluAlaGluThrAlaAlaValAlaTyrArgLysArgLeuMetThrGluGlyGly	261
PR	28-SEP-1999;	99US-0151551.	DB	782	AGGAGAAAGCTAAGAGATGCTGAGACGTGAAGAGAGAGTTTGAAGAGAGAGAGG	841
PR	28-SEP-1999;	99US-0151552.	QY	262	LeuAlaAlaAlaGluLysMetAspAlaArgGlyLeuLeuLeuValAlaCysPheGly	281
PR	28-SEP-1999;	99US-0151553.	DB	842	ATAGAGAAATGTGAAGACTCCTGATGTTTCATACGTTTTCAGCATCTTGTACATTGGA	901
PR	28-SEP-1999;	99US-0151554.	QY	282	ValProSerAsnPheArgSerThrAspLeuLeuAspLeuIleArgMetSerGlySerAsn	301
PR	28-SEP-1999;	99US-0151555.	DB	902	ATT-----GTTAAAGTGAAGATCTTGTGTTGTATAGAAAGCTT-----	940
PR	28-SEP-1999;	99US-0151556.	QY	302	GluIleAlaGlyAlaLeuLysArgSerGlnPhe-----Leu	313
PR	28-SEP-1999;	99US-0151557.	DB	941	---GTTCTGATCTCGTGGCGTAAACAGATGCTTAAGCTTTCGCTTCTCTGCTGTTG	997
PR	28-SEP-1999;	99US-0151558.	QY	314	ValProMetValSerGlyIleValGluSerSerIleLysArgGlyMetHisIleGluAla	333
PR	28-SEP-1999;	99US-0151559.	DB	998	GGTGATCAATGCTGACATGATTGAAGAAATTAATCAGCGGGGACACAGCTTGTATGCT	1057
PR	28-SEP-1999;	99US-0151560.	QY	334	LeuGluMetValTyrThrPheGlyMetGluAspLysPheSerAlaAlaLeuValLeuThr	353

Alignment Scores:
Pred. No.: 1.55e-14 Length: 1964


```
Db 1748 AGTCATCCCTGAAGG---TATCCGAGTATGCTATGACACACAGACCGTTCCTTAGCGG 1804
Qy 492 aLeuValSerTyrLeuGlyProSerThrSerPheProHisArgSerArgArgSerPi 512
Db 1805 TCCAATATGGCAGCACACACCTCCACCACTCTCTCTCTCAGACTTACACGTTCAATCC 1864
Qy 512 oGluTyrMetValProLeuProHisGlyGlyLeuGlyArgSerValTyrAlaTyrGluHi 532
Db 1865 T-----GTCCTT---GCATGGGAACTTACCGCACTGCTATCAGTACCGGC 1912
Qy 532 sLeuAlaProAnSerTyr 538
Db 1913 TCCTCTCTCTCTCTCTTAC 1931
RESULT 5
AAC49876
ID AAC49876 standard; DNA; 1964 BP.
XX
AC AAC49876;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 62772.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss
XX
OS Arabidopsis thaliana.
XX
FN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 03-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139484.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 19-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
```

PR 24-SEP-1999; 99US-0155659.
 PR 28-SEP-1999; 99US-0156458.
 PR 29-SEP-1999; 99US-0156596.
 PR 04-OCT-1999; 99US-0157117.
 PR 05-OCT-1999; 99US-0157753.
 PR 06-OCT-1999; 99US-0157865.
 PR 07-OCT-1999; 99US-0158029.
 PR 08-OCT-1999; 99US-0158232.
 PR 12-OCT-1999; 99US-0158369.
 PR 13-OCT-1999; 99US-0158393.
 PR 13-OCT-1999; 99US-0158294.
 PR 13-OCT-1999; 99US-0159295.
 PR 14-OCT-1999; 99US-0159329.
 PR 14-OCT-1999; 99US-0159330.
 PR 14-OCT-1999; 99US-0159331.
 PR 14-OCT-1999; 99US-0159637.
 PR 14-OCT-1999; 99US-0159638.
 PR 18-OCT-1999; 99US-0159584.
 PR 21-OCT-1999; 99US-0160741.
 PR 21-OCT-1999; 99US-0160767.
 PR 21-OCT-1999; 99US-0160768.
 PR 21-OCT-1999; 99US-0160770.
 PR 21-OCT-1999; 99US-0160814.
 PR 21-OCT-1999; 99US-0160815.
 PR 22-OCT-1999; 99US-0160880.
 PR 22-OCT-1999; 99US-0160981.
 PR 22-OCT-1999; 99US-0160989.
 PR 25-OCT-1999; 99US-0161404.
 PR 25-OCT-1999; 99US-0161405.
 PR 26-OCT-1999; 99US-0161359.
 PR 26-OCT-1999; 99US-0161360.
 PR 26-OCT-1999; 99US-0161361.
 PR 28-OCT-1999; 99US-0161920.
 PR 28-OCT-1999; 99US-0161992.
 PR 28-OCT-1999; 99US-0161993.
 PR 29-OCT-1999; 99US-0162142.

Alignment Scores:

Pred. No.: 4,848-18 Length: 2158
 Score: 340.00 Matches: 139
 Percent Similarity: 41.13% Conservatives: 86
 Best Local Similarity: 25.41% Mismatches: 217
 Query Match: 10.95% Indels: 105
 DB: 21 Gaps: 21

US-09-890-475-1 (1-609) x AAC46135 (1-2158)

QY 64 AlapheSerValAlaValGluThrPhe-LysArgGlnPhe-----AspAspLeuG1 80
 DB 392 GCATTTTCATGGCTTGAGAGGCTTTGAAGAGAGGTTTCATGAGCTTGAAGATCAAGA 451
 QY 80 nlyHisIleGluSerIleGluAsnAlaIleAspSerLysLeuGluSerAsnGlyValVa 100
 DB 452 AAAGAGATGAACCAAAACCAAGAAAGCTCAAGAACTGTTGGAGAAAAGAGGAGC 511
 QY 100 lleuAlalaArgAsn-----ProArgAsn-----AsnValSerValG1 123
 DB 512 TGTTCAGGCTAAGGAGAGGCGGCTTGAAGGCTTCAGAGAGAGAGAGACCGAGCTAT 571
 QY 106 -----AsnAsnPheIleGlnProMetLeuSerPr 115
 DB 572 GTTTACTATAACAGTGCTTTGGATAAGTATAACAAAC-----GCTCCTGTCAGTAAACC 625
 QY 115 O-----ProArgAsn-----AsnValSerValG1 123
 DB 626 TTCTGTTGGTGAACGCTTGCCACAAATGCTGTAGAGGATTCACCAATGCTTTGCTGC 685
 QY 123 uThrThrValThrValSerGlnPro-----SerGlnGluLeuValPro-- 137
 DB 686 TGATGATTATACCGATGATAATCCGGATGGTATTGTGCAAGATGTTTCAGATCTCACCTGT 745
 QY 138 -----GluThrSerAsnLysProGluGlyGlyArgMetCysGluLeuMetCy 153

DB 746 CATGGAAATATAGGTGAAGGCTTATCCACAGTTACTGAACACTGCGGGGATAGA 805
 QY 153 sSerLysGlyLeuArgLysTyrIleTyrAlaAsnIleSerAspGlnAlaLysLeuMetG1 173
 DB 806 CTCACGGGTCTCCATAAGTTTGTATCAGATAACCCCAAGAACCTTGCATCACTAAGA 865
 QY 173 uGluIleProSerAlaLeuLysLeuAlaLysGluProAlaLysPheValLeuAspCysI1 193
 DB 866 GGAGATTCCTATGGCGTTCAGGCTGCAACCAACCGAGCTTAGTGTGACTCTTT 925
 QY 193 eGlyLysPheTyr-----LeuGlnGlyArgAlaPheThrLysG1 207
 DB 926 GGAAGGGTTTTACCTATGAGGCACCAACTGCTGATGGAATA-----AAAGA 973
 QY 207 userProMetSerSerAlaArgGlnValSerLeuLeuLeu-----SerPheLe 225
 DB 974 CGCTAACCTCTTGGGCATGCGCAGAACCTGTATCATGTTGATGAGGTGCTTAGCATACT 1033
 QY 225 uLeuMetPro---AspArgGlyLysGlyLysValLysIleGluSerTrpIleLysAspG1 244
 DB 1034 GTTATCTGCTGCGACGCAACTGCTGCTGTGTTCTCTCACAAATGTTTAAGCATAG 1093
 QY 244 uAlaGluThrAlaAlaValAlaIlePargLysArgLeuMetThr---GluGlyGlyLeuAl 263
 DB 1094 AGCCAAACCTATTGCAAGAGGATGGAATCCACTGCTGGAATCTCTTGACATGGATGCTTG 1153
 QY 263 aAlaAlaGluLysMetAspAlaArgGlyLeuLeuLeuValAlaCysPheGlyValPr 283
 DB 1154 CAATGGTAATCTTGGAGGCTCATGCATCTCTGCAACTGCTGGCCACTTTTGTCTATTGT 1213
 QY 283 oSerAsnPheArgSerThrAspLeuLeuAspLeuIleArgMetSerGlySerAsnGluI1 303
 DB 1214 TGCTGATTTTAAAGAAAGATGAACCTCTTAAAGCTGATCCCACTGGTTTTCACGTCGCGTCA 1273
 QY 303 eAlaGlyAlaLeuLysArgSerGlnPheLeuValProMetValSerGlyLeuValGluSe 323
 DB 1274 GCGAGCTGAGCTCTGCGCTTCACTGATTAGCTGAAAAAATGCCGTGGTGTGATGAAGT 1333
 QY 323 rSerIleLysArgGlyMetHisIleGluAlaLeuGluMetValTyrThrPheGlyMetG1 343
 DB 1334 TCTGCTGAACAGCTGGAACAGATGATGCGGTAACTTGGCATTTTGGCTTTTGAACCTCAC 1393
 QY 343 uAspLysPheSerAlaAlaLeuValLeuThrSerPheLeuLysMetSerLysGluSerPh 363
 DB 1394 AGAACCTTCTCACCTGCTCATTAATCTTACTTAGTAGAGGCAAGCGATCTTC 1453
 QY 363 eGluArgAlaLysArgLysAlaGlnSerProLeu-----AlaPheLysGluAlaAlaTh 381
 DB 1454 CCCCCAAGGCAGACCTGGCAATGCATCTCTGCTGTTCAGGCAAGTATGATGATCAATGA 1513
 QY 381 rLysGluLeuAlaValLeuSerSerValMetGlnCysMetGluThrHisLysLeuAspPr 401
 DB 1514 GCGGAGCTTATAGGCTCAAAACCGTTATAAGTGTATCGAAGAGCATAGGCTGGAA-- 1571
 QY 401 oAlaLeuGluLeuProGlyTrpGlnIleLysGluGlnIleValSerLeuGluLysAspTh 421
 DB 1572 ----GACGAGTACCCA-----GTTGAGCCACTTCACAAACGGAT 1606
 QY 421 rLeuGlnLeuAspLysGluMetGluGluLysAlaArgSerLeuSerLeuMetGluGluAl 441
 DB 1607 TCTCCAGCTCGAAGAAACCCAAAGCGGACAAAGAGAGAGCAACAGAACCCCATGAAA-- 1661
 QY 441 aAlaLeuAlaLysArgMetTyrAsnGlnGlnIleLysArgProArgLeuSerProMetG1 461
 DB 1662 -----CCTCAGCCAAACCGACCGTGGTGTCT----- 1688
 QY 461 uMetProProValThrSerSerSer-----TyrSerPr 472
 DB 1689 -CAACCCGAGTCACTGACAAACAAACAAACAACTCAACAAACAAAGACAGGATAGGTAG 1747
 QY 472 oIleTyrArgAspArgSerPheProSerGlnArgAspAspGlnAspGluIleSerAl 492

XX	25-FEB-1999;	99US-0121825.
PF	25-FEB-1999;	99US-0121825.
XX	05-MAR-1999;	99US-0123180.
PR	05-MAR-1999;	99US-0123180.
PR	09-MAR-1999;	99US-0123548.
PR	23-MAR-1999;	99US-0125788.
PR	23-MAR-1999;	99US-0126264.
PR	25-MAR-1999;	99US-0126785.
PR	29-MAR-1999;	99US-0126785.
PR	01-APR-1999;	99US-0127462.
PR	06-APR-1999;	99US-0128234.
PR	08-APR-1999;	99US-0128714.
PR	16-APR-1999;	99US-0129845.
PR	19-APR-1999;	99US-0130077.
PR	21-APR-1999;	99US-0130449.
PR	23-APR-1999;	99US-0130510.
PR	23-APR-1999;	99US-0130891.
PR	28-APR-1999;	99US-0131149.
PR	30-APR-1999;	99US-0132048.
PR	30-APR-1999;	99US-0132407.
PR	04-MAY-1999;	99US-0132484.
PR	05-MAY-1999;	99US-0132485.
PR	06-MAY-1999;	99US-0132486.
PR	06-MAY-1999;	99US-0132487.
PR	07-MAY-1999;	99US-0132863.
PR	11-MAY-1999;	99US-0134256.
PR	14-MAY-1999;	99US-0134218.
PR	14-MAY-1999;	99US-0134219.
PR	14-MAY-1999;	99US-0134221.
PR	14-MAY-1999;	99US-0134370.
PR	18-MAY-1999;	99US-0134768.
PR	19-MAY-1999;	99US-0134941.
PR	20-MAY-1999;	99US-0135124.
PR	21-MAY-1999;	99US-0135353.
PR	24-MAY-1999;	99US-0135629.
PR	25-MAY-1999;	99US-0136021.
PR	27-MAY-1999;	99US-0136392.
PR	28-MAY-1999;	99US-0136782.
PR	01-JUN-1999;	99US-0137222.
PR	03-JUN-1999;	99US-0137528.
PR	04-JUN-1999;	99US-0137502.
PR	07-JUN-1999;	99US-0137724.
PR	08-JUN-1999;	99US-0138094.
PR	10-JUN-1999;	99US-0138540.
PR	10-JUN-1999;	99US-0138647.
PR	14-JUN-1999;	99US-0139119.
PR	16-JUN-1999;	99US-0139452.
PR	17-JUN-1999;	99US-0139453.
PR	18-JUN-1999;	99US-0139492.
PR	18-JUN-1999;	99US-0139454.
PR	18-JUN-1999;	99US-0139455.
PR	18-JUN-1999;	99US-0139456.
PR	18-JUN-1999;	99US-0139457.
PR	18-JUN-1999;	99US-0139458.
PR	18-JUN-1999;	99US-0139459.
PR	18-JUN-1999;	99US-0139460.
PR	18-JUN-1999;	99US-0139461.
PR	18-JUN-1999;	99US-0139462.
PR	18-JUN-1999;	99US-0139463.
PR	18-JUN-1999;	99US-0139750.
PR	21-JUN-1999;	99US-0139763.
PR	21-JUN-1999;	99US-0139817.
PR	22-JUN-1999;	99US-0139899.
PR	23-JUN-1999;	99US-0140353.
PR	23-JUN-1999;	99US-0140354.
PR	24-JUN-1999;	99US-0140695.
PR	24-JUN-1999;	99US-0140823.
PR	24-JUN-1999;	99US-0140891.
PR	30-JUN-1999;	99US-0141287.
PR	01-JUL-1999;	99US-0141842.
PR	01-JUL-1999;	99US-0142154.
PR	02-JUL-1999;	99US-0142055.
PR	06-JUL-1999;	99US-0142350.
PR	08-JUL-1999;	99US-0142803.
PR	09-JUL-1999;	99US-0142920.
PR	12-JUL-1999;	99US-0142977.
PR	13-JUL-1999;	99US-0143542.
PR	14-JUL-1999;	99US-0143624.
PR	15-JUL-1999;	99US-0144005.
PR	16-JUL-1999;	99US-0144085.
PR	16-JUL-1999;	99US-0144086.
PR	19-JUL-1999;	99US-0144325.
PR	19-JUL-1999;	99US-0144331.
PR	19-JUL-1999;	99US-0144332.
PR	19-JUL-1999;	99US-0144333.
PR	19-JUL-1999;	99US-0144334.
PR	19-JUL-1999;	99US-0144335.
PR	20-JUL-1999;	99US-0144352.
PR	20-JUL-1999;	99US-0144632.
PR	20-JUL-1999;	99US-0144884.
PR	21-JUL-1999;	99US-0144814.
PR	21-JUL-1999;	99US-0145086.
PR	21-JUL-1999;	99US-0145088.
PR	22-JUL-1999;	99US-0145085.
PR	22-JUL-1999;	99US-0145087.
PR	22-JUL-1999;	99US-0145089.
PR	22-JUL-1999;	99US-0145192.
PR	23-JUL-1999;	99US-0145145.
PR	23-JUL-1999;	99US-0145218.
PR	23-JUL-1999;	99US-0145224.
PR	26-JUL-1999;	99US-0145276.
PR	27-JUL-1999;	99US-0145913.
PR	27-JUL-1999;	99US-0145918.
PR	27-JUL-1999;	99US-0145919.
PR	28-JUL-1999;	99US-0145951.
PR	02-AUG-1999;	99US-0146386.
PR	02-AUG-1999;	99US-0146388.
PR	02-AUG-1999;	99US-0146389.
PR	03-AUG-1999;	99US-0147038.
PR	04-AUG-1999;	99US-0147204.
PR	04-AUG-1999;	99US-0147302.
PR	05-AUG-1999;	99US-0147192.
PR	05-AUG-1999;	99US-0147280.
PR	06-AUG-1999;	99US-0147303.
PR	06-AUG-1999;	99US-0147416.
PR	09-AUG-1999;	99US-0147493.
PR	09-AUG-1999;	99US-0147935.
PR	10-AUG-1999;	99US-0148171.
PR	11-AUG-1999;	99US-0148319.
PR	12-AUG-1999;	99US-0148341.
PR	13-AUG-1999;	99US-0148565.
PR	13-AUG-1999;	99US-0148684.
PR	16-AUG-1999;	99US-0149368.
PR	17-AUG-1999;	99US-0149175.
PR	18-AUG-1999;	99US-0149426.
PR	20-AUG-1999;	99US-0149722.
PR	20-AUG-1999;	99US-0149723.
PR	20-AUG-1999;	99US-0149922.
PR	23-AUG-1999;	99US-0149902.
PR	23-AUG-1999;	99US-0149930.
PR	25-AUG-1999;	99US-0150566.
PR	25-AUG-1999;	99US-0150884.
PR	27-AUG-1999;	99US-0151065.
PR	27-AUG-1999;	99US-0151066.
PR	27-AUG-1999;	99US-0151080.
PR	30-AUG-1999;	99US-0151303.
PR	31-AUG-1999;	99US-0151438.
PR	01-SEP-1999;	99US-0151930.
PR	07-SEP-1999;	99US-0152363.
PR	10-SEP-1999;	99US-0153070.
PR	13-SEP-1999;	99US-0153758.
PR	15-SEP-1999;	99US-0154018.
PR	16-SEP-1999;	99US-0154039.
PR	20-SEP-1999;	99US-0154779.
PR	22-SEP-1999;	99US-0155139.
PR	23-SEP-1999;	99US-0155486.

PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139422.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 21-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147132.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147433.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149358.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 23-AUG-1999; 99US-0149930.
PR 23-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151338.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 13-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160988.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.

Db 1474 AATGAGATTCCGGTGCTTTGAAGCGGTACAGTTTCTTGTCCCTATGGTCTCAGGTACC 1533
Qy 318 ----- 318
Db 1534 ATATTCGTGTTCTCACTCGGTGAATTTCAATGCAAGGTGGTTCCTTTGTTGGACATCATC 1593
Qy 318 ----- 318
Db 1594 GACCAACATCAAGTTCATCTTTGTTTTCGATAAGCTTGATGGTATATAAATAGGAGAGC 1653
Qy 318 ----- 318
Db 1654 ACATCAATATTTAGAGTGCAATGACTGATGAGCCAAATCTAGCTAGAAATTAATCTG 1713
Qy 318 ----- 318
Db 1714 GAAAGAACTTGGAACTCTCAACCATAGTTTGGTAGCAAAATTTGTTGTTGTCAGAACCA 1773
Qy 318 ----- 318
Db 1774 AATGATAGGCTATTGCCCTTGAATAGTGTTCCTTGTGGTTTCCAAATATTGGAAGTTAAAA 1833
Qy 318 ----- 318
Db 1834 TCGTATGACTTAGCTGTGGTACTAATTAAGCTTAAGCAATGCCAACTCTAAGAAGTGG 1893
Qy 319 ----- GlyIleValGluSerSerIleLysArgGlyMet 329
Db 1894 TACTTACACATATTCATTATGGTCATAGGTATAGTTGATCATAGTATCAAGCGTGAATG 1953
Qy 330 HistGluAlaLeuGluMetValTyrPheGlyMetGluAspLysPheSerAlaIala 349
Db 1954 CATATTGAAGCTCTTGAGATGGTTTATACCTTTGGCATGGAGATAAGTTTTCAGCTGCT 2013
Qy 350 LeuValLeuThrSerPheLeuLysMetSerLysGluSerPheGluArgAlaLysArgLys 369
Db 2014 CTAGTTCTTAACCTTCATTCTTAAAGTAGAGCAAGAGTCATTTGAGAGGCGCAAAACGGAAA 2073
Qy 370 AlaGlnSerProLeuAlaPheLys----- 377
Db 2074 GCCAGTCCACCGCTGGCATTTGT-ATGAACCCCTTCCTTGCACATTTATGTACCTTTATGA 2132
Qy 378 ----- GluAla 379
Db 2133 ACTCTTTATCATCATCTGAGTCTGACCATGATGATATTTATTTCTCAACAGAAAGAGCG 2192
Qy 380 AlaThrLysGlnLeuAlaValLeuSerSerValMetGlnCysMetGluThrHisLysLeu 399
Db 2193 GCTACAAAGCAGCTAGCTGTGTATCATCAGTTATGCAGTGTATGGAGACTCACAAAGTTA 2252
Qy 400 AspProAlaLysGluLeuProGlyTyrGlnIleLysGluGlnIleValSerLeuGlyLys 419
Db 2253 GATCCTGCGAAGAACTACCCAGGATGGCAGATCAAGAGCAAAATTTGTAGCTTGGAGAAA 2312
Qy 420 AspThrLeuGlnLeuAspLysGluMetGluLysAlaArgSerLeuSerLeuMetGlu 439
Db 2313 GACACTCTTCAGCTCGACAAAGAGATGGAAGAGAAAGAGATCTCTCAGTTTAATGGAG 2372
Qy 440 GluAlaAlaLeuAlaLysArgMetTyrAsnGlnIleLysArgProArgLeuSerPro 459
Db 2373 GAAGCCGCACTTGCCAAAGAGAATGTATAACCAACAGATAAAACGTCACAGGTTGTACCC 2432
Qy 460 MetGluMetProProValThrSerSerTyrSerProIleTyrArgAspArgSerPhe 479
Db 2433 ATGGAATGCCACCAAGTAACTCTTCATCGTATTTCTCTATCTACCGTGTAGAGCTTT 2492
Qy 480 ProSerGlnArgAspAspGlnAspGluIleSerAlaLeuValSerSerTyrLeuGly 499
Db 2493 CCTAGTCAAGAGAGCAGTACCAAGATGAAATATCAGCTCTTTGTAGTAGTTACCTCGGC 2552
Qy 500 ProSerThrSerPheProHisArgSerArgSerProGluTyrMetValProLeuPro 519
Db 2553 CCGTCAACATCTTTTCTCTCATCGCTCAAGAGATCCCCCGGAATATATGTTTCCACTTCCA 2612

Qy 520 HisGlyGlyLeuGlyArgSerValTyrAlaTyrGluHisLeuAlaProAsnSerTyrSer 539
Db 2613 CATGGTGGGTTTAGGAAGAAGTATATCATATGACATCTGGCCCCCAATTCATCTCT 2672
Qy 540 ProGlyHisGlyHisArgLeuHisArgGlnTyrSerProSerLeuValHisGlyGlnArg 559
Db 2673 CCAGGTCCAGGACATAGACTTCATCGACAGTACTCTCGTCTTTGGTTTCAGGACAGAGA 2732
Qy 560 HisProLeuGlnTyrSerProProIleHisGlyGlnGlnLeuProTyrGlyIleGln 579
Db 2733 CATCCACTACAGTACTCTCTCCATTCATGACAAACACAGTTACCATATGGTATACAA 2792
Qy 580 ArgValTyrArgHisSerProSerGluGluArgTyrLeuGlyLeuSerAsnGlnArgSer 599
Db 2793 AGGGTTTACAGACATTCACCATCTCAAGAAAGATATTTGGGTTTATCCAAATCAAGGTCT 2852
Qy 600 ProArgSerAsnSerSerLeuAspProLys 609
Db 2853 CCTCGCAGTAACATCATTCATTAGACCCCAAA 2892
RESULT 3
AAC33344
ID AAC33344 standard; DNA; 2155 BP.
XX AAC33344;
AC AAC33344;
DT 17-OCT-2000 (first entry)
XX Arabidopsis thaliana DNA fragment SEQ ID NO: 2690.
XX Hybridisation assay; genetic mapping; gene expression control;
XX protein identification; signal transduction pathway;
XX metabolic pathway; promoter; termination sequence; ss.
XX Arabidopsis thaliana.
PN EP1033405-A2.
XX 06-SEP-2000.
PF 25-FEB-2000; 2000EP-0301439.
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125768.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.

QY 501 SerThrSerPheProHisArgSerArgSerProGluTyrMetValProLeuProHis 520
 Db 1862 TCAACATCTTTTCCATCGCTCAAGAGATCCCGGAATATATGGTCCACATCCACAT 1921
 QY 521 GlyGlyLeuGlyArgSerValTyrAlaTyrGluHisLeuAlaProAnSerTyrSerPro 540
 Db 1922 GGTGGGTTAGGAAGAAGTGATATGTCATATGAACATCTGGCCCCAAATTCATCTCTCCA 1981
 QY 541 GlyHisGlyHisArgLeuHisArgGlnTyrSerProSerLeuValHisGlyGlnArgHis 560
 Db 1982 GGTCCAGGACATAGACTTCATCGACAGTACTCTCCGCTTTGGTTCACGACAGACAT 2041
 QY 561 ProLeuGlnTyrSerProProIleHisGlyGlnGlnLeuProTyrGlyIleGlnArg 580
 Db 2042 CCACCTACAGTACTCTCTCCCAATTCATGGACAACAACAGITACCATATGTTATACAAAG 2101
 QY 581 ValTyrArgHisSerProSerGluGluArgTyrLeuGlyLeuSerAnGlnArgSerPro 600
 Db 2102 GTTTACAGACATTCACCATCTGAAGAAGATATTTGGGTTTATCCAAATCAAGGTCTCCT 2161
 QY 601 ArgSerAnSerSerLeuAspProllys 609
 Db 2162 GCAGTAATCTATCATTTAGACCCCAAA 2188

RESULT 2

ID AAA63668 standard; DNA; 3761 BP.

AC AAA63668;

XX 04-DEC-2000 (first entry)

DT 04-DEC-2000 (first entry)

DE H51 genomic sequence of first 17 kb of 84M13 encompassing FRI gene.
 KW H51; one locus-FRIGIDA; FRI gene; flowering time; blotting;
 KW flower initiation; stem elongation; flower production; ss.

XX Arabidopsis sp.

XX WO200046359-A2.

XX 10-AUG-2000.

PF 25-JAN-2000; 2000WO-GB00197.

PR 05-FEB-1999; 99GB-0002660.

XX (PLAN-) PLANT BIOSCIENCE LTD.

PI Johanson U, West J, Dean C;

XX WPI; 2000-532899/48.

XX New nucleic acid derived from the FRI locus of a plant, e.g.

PT Arabidopsis, encoding a polypeptide capable of specifically altering

PT the flowering time of a plant.

XX Claim 4; Fig 4; 73pp; English.

BS The present sequence represents the (late flowering) H51 genomic

CC sequence of the first 17 kb of cosmid 84M13 encompassing FRI (one

CC locus-FRIGIDA) locus of Arabidopsis. The FRI gene encodes a

CC polypeptide capable of specifically altering the flowering time of a

CC plant. The FRI polynucleotide is used to transform plants, so that

CC the flowering time of a plant is altered. This is used, for example,

CC for plants in which the leaves or tubers are a commercial product,

CC where it is desirable to avoid 'blotting' (initiation of flowers and

CC stem elongation) at too early a stage. Conversely, it may be desirable

CC to alter flowering under certain circumstances e.g. to vary flower

CC production across the seasons.

XX Sequence 3761 BP; 1174 A; 717 C; 775 G; 1095 T; 0 other;

Alignment Scores:
 Pred. No.: 3761
 Score: 78.96%
 Percent Similarity: 78.96%
 Best Local Similarity: 78.96%
 Query Match: 162
 DB: 21
 Gaps: 2

US-09-890-475-1 (1-609) x AAA63668 (1-3761)

QY 1 MetSerAnTyrProThrValAlaAlaGlnProThrThrAlaAnProLeuLeu 20
 Db 574 ATGTCCCAATTATCCACGACGGTGGCGGCAACCCACACGCGCAATCCACTGCTG 633
 QY 21 GlnArgHisGlnSerGluGlnArgArgGluLeuProLysIleValGluThrGluSer 40
 Db 634 CAGCGACATCAATCTGAACAGCGACGAGAGAAATTAACGAGATTGTCGAAACAGAGTCT 693
 QY 41 ThrSerMetAspIleThrIleGlyGlnSerLysGlnProGlnPheLeuLysSerIleAsp 60
 Db 694 ACAAGTATGGACATTACGATCGGTCAATCTAAGCAGCCTCAATTTTGAATCCATAGAC 753
 QY 61 GluLeuAlaAlaPheSerValAlaValGluThrPheLysArgGlnPheAspLeuGln 80
 Db 754 GAATTAGCTCGTTTTTCAGTTGCAGTGGAAACATTCAAACGCCAATTCGATGATCTTCAG 813
 QY 81 LysHisIleGluSerIleGluAsnAlaIleAspSerLysLeuGluSerAnGlyValVal 100
 Db 814 AAGCACATCGAGTCAATCGAAAACGCAATTGATTCCAACTCGAGAGTAACGGCGTTGTC 873
 QY 101 LeuAlaAlaArgAnAnAnPheHisGlnProMetLeuSerProProArgAnAnVal 120
 Db 874 CTCGCCCGCGGGAACAATAATTTCCATCAGCCGATGTTATCGCCTCCGCGGAACAATGTA 933
 QY 121 SerValGluThrThrValThrValSerGlnProSerGlnGluIleValProGluThrSer 140
 Db 934 TCTGTAGAACCCACCGTCACTGTGAGCCAAACCGTCTCAGAGAGATTGTACGGAGCGTGC 993
 QY 141 AsnLysProGluGlyArgMetCysGluLeuMetCysSerLysGlyLeuArgLysTyr 160
 Db 994 AATAAACCGGAGGGGACGATGTGTGTGAGTTGATGTAGCAAAAGGTCTGGCTAAATAC 1053
 QY 161 IleTyrAlaAnIleSerAspGlnAlaLysLeuMetGluGluIleProSerAlaLeuLys 180
 Db 1054 ATATACCGGAATATCTCTGATCAAGCTAAGTTAATGGAAGAGATTCCTTCAGCTTGAAA 1113
 QY 181 LeuAlaLysGluProAlaLysPheValLeuAspCysIleGlyLysPheTyrLeuGlnGly 200
 Db 1114 TTGGCCAAGGAGCGAGCAAGTTTGTATTGGATTGTATTGGCAAGTTTACTTACAAGGG 1173
 QY 201 ArgAlaAlaPheThrLysGluSerProMetSerSerAlaArgGlnValSerLeuLeuIle 220
 Db 1174 CGTAGAGCATTTACTAAAGAGTCGCTATGAGTCTGCGAGACAAGTTTCCGTTCTTTATA 1233
 QY 221 LeuGluSerPheLeuLeuMetProAspArgGlyLysGlyLysValLysIleGluSerTyr 240
 Db 1234 CTGGAGTCTTTTCTCTAAATGCTGATCGTGGTAAAGGGAAGGTGAAGATTGAGAGTTGG 1293
 QY 241 IleLysAspGluAlaGluThrAlaAlaValAlaTyrArgLysArgLeuMetThrGluGly 260
 Db 1294 ATTAAGATGAGCGGAGACGGTGTGTTGCTGTGGAGAAAAGGTGTGATGACTGAAGA 1353
 QY 261 GlyLeuAlaAlaAlaGluLysMetAspAlaAtgGlyLeuLeuLeuValAlaCysPhe 280
 Db 1354 GGATTAGTCGGCTGAGAAAATGGATGCAAGGGGTTTGTCTTTACTAGTTGCTGTTT 1413
 QY 281 GlyValProSerAnPheArgSerThrAspLeuLeuAspLeuIleArgMetSerGlySer 300
 Db 1414 GGTCTTCTCTCAAACTTTAGAGTACAGATTTCCTGGATTGATAGAGGATGAGTGGTTCG 1473
 QY 301 AsnGluIleAlaGlyAlaLeuLysArgSerGlnPheLeuValProMetValSer----- 318

FT /*tag= a
 FT /product= "polypeptide which alters flowering time"
 XX WO200046358-A2.
 XX 10-AUG-2000.
 XX 25-JAN-2000; 2000WO-GB00197.
 XX 05-FEB-1999; 99GB-0002660.
 PR (PLAN-) PLANT BIOSCIENCE LTD.
 XX Johanson U, West J, Dean C;
 XX WPI; 2000-532899/48.
 DR P-PSDB; AAB08030.
 XX New nucleic acid derived from the FRI locus of a plant, e.g.
 FT Arabidopsis, encoding a polypeptide capable of specifically altering
 PT the flowering time of a plant
 XX Claim 4; Fig 5; 73pp; English.

CC The present sequence encodes a polypeptide capable of specifically
 CC altering the flowering time of a plant. The polypeptide is encoded
 CC by the FRI (one locus-FRIGIDA) locus of Arabidopsis. The FRI
 CC polynucleotide is used to transform plants, so that the flowering time
 CC of a plant is altered. This is used, for example, for plants in which
 CC the leaves or tubers are a commercial product, where it is desirable
 CC to avoid 'blotting' (initiation of flowers and stem elongation) at
 CC too early a stage. Conversely, it may be desirable to alter flowering
 CC under certain circumstances e.g. to vary flower production across the
 CC seasons.

SQ Sequence 2257 BP; 698 A; 443 C; 493 G; 623 T; 0 other;

Alignment Scores:

Pred. No.: 2,96e-244 Length: 2257
 Score: 3104.00 Matches: 609
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatch: 0
 Query Match: 100.00% Indels: 0
 DB: 21 Gaps: 0

US-09-890-475-1 (1-609) x AAA63669 (1-2257)

Qy 1 MetSerAsnTyrProThrValAlaAlaGlnProThrThrAlaAsnProLeuLeu 20
 Db 362 ATGTCCTAATTATCACCGACGGTGGCGGCGCAACCCACACGACGCGAATCCACTGCTG 421
 Qy 21 GlnArgHisGlnSerGluGlnArgArgGluLeuProIysIleValGluThrGluSer 40
 Db 422 CAGCGACATCAATCTGAACGACGACGACGACGACGACGACGACGACGACGACGACGACG 481
 Qy 41 ThrSerMetAspIleThrIleGlyGlnSerIysGlnProGlnPheLeuIysSerIleAsp 60
 Db 482 ACAAGTATGACATTCAGTCGCTCAATCTAAGCAGCCTCAATTTTGAATCCATAGAC 541
 Qy 61 GluLeuAlaAlaPheSerValAlaValGluThrPheIysArgGlnPheAspIleGln 80
 Db 542 GAATTAGTCGCTTTTTCAGTTGAGTGGAAACATTCAAACGCGCAATTCGATGATCTTCAG 601
 Qy 81 LysHisIleGluSerIleGluAsnAlaIleAspSerIysLeuGluSerAsnGlyValVal 100
 Db 602 AAGCACATCGAGTCATTCGAACGCAATTCGATTCGAACGAGTGAACGCGGTGTC 661
 Qy 101 LeuAlaAlaArgAsnAsnAsnPheHisGlnProMetIleSerProProArgAsnVal 120
 Db 662 CTCGCGCGCGGGAACAATAATTCATTCAGCCGATGTTATCGCTCCGCGGAACAATGTA 721
 Qy 121 SerValGluThrThrValThrValSerGlnProSerGlnGluIleValProGluThrSer 140

Db 722 TCTGTAGAAACACACCGTCACTGTGAGCCACCGTCTCAGGAGATTGTACCGGAGACGTGC 781
 Qy 141 AsnIysProGluGlyArgMetCysGluLeuMetCysSerIysGlyLeuArgIysTyr 160
 Db 782 AATAAACCGGAGGGGAGCGTATGTGTAGTTGATGTAGCAAAAGGTCTCGTAAATAC 841
 Qy 161 IleTyrAlaAsnIleSerAspGlnAlaIysLeuMetGluGluIleProSerAlaLeuIys 180
 Db 842 ATATACCGAATATCTCTGATCAAGCTAAGTTAATGAGAGAGATTCCTTCAGCTTTGAA 901
 Qy 181 LeuAlaIysGluProAlaIysPheValLeuAspCysIleGlyIysPheTyrLeuGlnGly 200
 Db 902 TTGGCCAGGAGCGCAAGATTTCATTGGATTGTTATGGCAAGTTTACTTCAAGGG 961
 Qy 201 ArgArgAlaPheThrIysGluSerProMetSerSerAlaArgGlnValSerLeuLeu 220
 Db 962 CGTAGAGCAATTTACTAAGAGTCGCTATGAGCTCTCGAGACAGATTTCGCTTCTATA 1021
 Qy 221 LeuGluSerPheLeuLeuMetProAspArgGlyIysGlyIysValIysIleGluSerTyr 240
 Db 1022 CTGGAGTCTTTTCTTAATGCTGATCGTGTAAAGGGAAGGTCAAGATTGAGAGTTGG 1081
 Qy 241 IleIysAspGluAlaGluThrAlaAlaValAlaIleArgIysArgLeuMetThrGluGly 260
 Db 1082 ATTAAGATGAGCGGAGACGCTCTGCTGCTGGAGGAAAAGTTGATGACTGAAGGA 1141
 Qy 261 GlyIleAlaAlaAlaGluIysMetAspAlaArgGlyLeuLeuLeuValAlaCysPhe 280
 Db 1142 GGATTAGTCGCGCTGAGAAAATGATGATGCAAGGGGTTTGTCTTTACTAGTTGCTGTTT 1201
 Qy 281 GlyValProSerAsnPheArgSerThrAspIleuAspIleuAspIleuAspIleuAsp 300
 Db 1202 GGTGTTCTCTTCAAACTTAGGAGTACAGATTGCTGGATTGATAAGGATGAGTGGTTCG 1261
 Qy 301 AsnGluIleAlaGlyAlaLeuIysArgSerGlnPheLeuValProMetValSerGlyIle 320
 Db 1262 AATGAGATTGCGGTCGTTTGAAGCGGTCAAGTTTCTTGTCCCTATGCTCAGGTATA 1321
 Qy 321 ValGluSerSerIleIysArgGlyMetHisIleGluAlaLeuGluMetValTyrPhe 340
 Db 1322 GTTGAATCAAGTATCAAGCGTGAATGATGATTTGAAGCTCTTGAGATGGTTATACCTTT 1381
 Qy 341 GlyMetGluAspIysPheSerAlaAlaLeuValLeuThrSerPheLeuIysMetSerIys 360
 Db 1382 GGCATGGAGGATTAAGTTTTCAGTCTCTAGTTCTAATCTTCTTAAAGATGAGCAAG 1441
 Qy 361 GluSerPheGluArgAlaIysArgIysAlaGlnSerProLeuAlaPheIysGluAlaAla 380
 Db 1442 GAGTCATTTGAGAGGGGCAAAACGGAAGCCAGTCAACGCTGGCATTTTAAAGAGCGGCT 1501
 Qy 381 ThrIysGlnLeuAlaValLeuSerSerValMetGlnCysMetGluThrHisIysLeuAsp 400
 Db 1502 ACAAGCAGCTAGCTGTGTTATCATCATGTTATGAGTGTATGGAGACTCACAGTTAGAT 1561
 Qy 401 ProAlaIysGluLeuProGlyTyrGlnIleIysGluGlnIleValSerLeuGluIysAsp 420
 Db 1562 CCTGCGAAAGAACTACAGATGGCAGATCAAGAGCAAAATGTTAGCTTGGAGAAAGAC 1621
 Qy 421 ThrIleuGluIleuAspIysGluMetGluIysAlaArgSerLeuSerLeuGluGlu 440
 Db 1622 ACTCTTCAGCTCGCAAGAGATGGAAGAGAAAGCAAGATCTCTCAGTTTAAATGGAGAA 1681
 Qy 441 AlaAlaLeuAlaIysArgMetTyrAsnGlnGlnIleIysArgProArgLeuSerProMet 460
 Db 1682 CGCGCACTTGCACAGAGAAATGTATAACCAACAGATAAAACGTCACCAAGGTTGTACCCTATG 1741
 Qy 461 GluMetProProValThrSerSerSerTyrSerProIleTyrArgAspArgSerPhePro 480
 Db 1742 GAAATCCCAAGTAACTTCTTCATCTGATTTCTCTATCTACCGGTAGTAGAGGTTTCTCT 1801
 Qy 481 SerGlnArgAspAspGlnAspGluIleSerAlaLeuValSerSerTyrLeuGlyPro 500
 Db 1802 AGTCAAGAGACGATGACCAAGATGAATATACAGTCTTGTGAGTAGTATTACTCGGCCCG 1861

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: August 25, 2003, 08:46:35 ; Search time 462 Seconds
(without alignments)
3558.350 Million cell updates/sec

Title: US-09-890-475-1

Perfect score: 3104

Sequence: 1 MSNYPPTVAAGPTTANPLL.....RYLGSLNQRSPRNSSLDPK 609

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame_p2n.model -DEV=xlh
-Q=/cgn2.1/USPTO.spool/US09890475/runat.14082003.090348.19534/app_query.fasta_1.775
-DB=N_Geneseq.19jun03 -QFMT=fastcap -SUFFIX=ring -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS-human40.cdi
-LIST=45 -DOCLIGN=200 -THR_SCORE=Pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pcp -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09890475@cgn.1.1.312 @runat.14082003.090348.19534 -NCPU=6 -ICPU=3
-NO_MMWP -LARGESQUE -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N_Geneseq.19jun03.*

1: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*
25: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed,

and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	3104	100.0	2257	21	CDNA sequence of t
2	2981	96.0	3761	21	H51 genomic sequen
3	349	11.2	2155	21	Arabidopsis thalia
4	340	11.0	2158	21	Arabidopsis thalia
5	296.5	9.6	1964	21	Arabidopsis thalia
6	296.5	9.6	1970	21	Arabidopsis thalia
7	288.5	9.3	1865	21	Arabidopsis thalia
8	234.5	7.6	1524	24	Arabidopsis thalia
9	152	4.9	7785	23	Arabidopsis thalia
10	151	4.9	3312	20	Drosophila melanog
11	148.5	4.8	3824	20	Arabidopsis thalia
12	148.5	4.8	4620	25	Breast cancer asso
13	148.5	4.8	4972	24	Novel human coding
14	145.5	4.7	2760	24	DNA encoding human
15	145	4.7	2187	24	Human novel polynu
16	145	4.7	2045	24	Human DNA sequence
17	144	4.6	2684	24	Human novel polynu
18	144	4.6	2756	24	Human lung cancer
19	143	4.6	3270	25	Human intracellula
20	143	4.6	18506	23	Drosophila melanog
21	142	4.6	3114	18	Human RHAMM cDNA
22	142	4.6	3114	24	Human cDNA encodin
23	141.5	4.6	4772	23	Drosophila melanog
24	141	4.5	2020	23	Human novel polynu
25	140.5	4.5	2511	22	Human polynucleoti
26	140	4.5	3718	24	Rat sequence diffe
27	137.5	4.4	6558	21	CDNA sequence of a
28	136	4.4	6455	24	Human cDNA differe
29	136	4.4	7080	23	DNA encoding novel
30	135.5	4.4	14278	23	Drosophila melanog
31	135.5	4.4	14278	23	Drosophila melanog
32	134	4.3	2526	24	Arabidopsis thalia
33	134	4.3	11528	23	Arabidopsis thalia
34	133	4.3	17612	19	Pseudomonas Xpc, O
35	133	4.3	17612	21	Pseudomonas alcali
36	133	4.3	17612	22	Pseudomonas alcali
37	133	4.3	17612	24	Pseudomonas alcali
38	132.5	4.3	3913	22	Human polynucleoti
39	132.5	4.3	5142	24	CDNA encoding huma
40	132	4.3	7903	23	Arabidopsis thalia
41	131.5	4.2	2904	23	Drosophila melanog
42	131	4.2	9775	20	N. crassa his-3 co
43	130.5	4.2	3397	22	Human polynucleoti
44	130.5	4.2	7393	22	Human polynucleoti
45	130	4.2	6949	23	Drosophila melanog

ALIGNMENTS

RESULT 1

AAA63669

ID AAA63669 standard; cDNA; 2257 BP.

XX AAA63669;

AC AAA63669;

XX 04-DEC-2000 (first entry)

DT CDNA sequence of the FRI gene of Arabidopsis.

XX H51; one locus-FRIGIDA; FRI gene; flowering time; blotting;

KW flower initiation; stem elongation; flower production; ss.

XX Arabidopsis sp.

XX Key

XX Location/Qualifiers

XX 362...2191

DIQKHLESTENAIKSLKESNGVLAARNNNHOPMLSPENNUSVETTVTSQPSQEI
VPETSNKPEGCMELMSKGLRKYIYANISDQAKLMEBIPSAKLAKPAKFVLDCl
GAFYLGRRRAFTKESFMSARQVSLI"
BASE COUNT 166 a 115 c 133 g 144 t
ORIGIN

Alignment Scores:
Pred. No.: 6.5e-53 Length: 558
Score: 931.00 Matches: 185
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.48% Mismatches: 0
Query Match: 29.99% Indels: 0
DB: 8 Gaps: 0

US-09-890-475-1 (1-609) x AY092693 (1-558)

QY	34	LysIleValGluThrGluSerThrSerMetAspIleThrIleGlyGlnSerLysGlnPro	53
Db	1	AAAGATTGCGAAACAGAGTCTACAGATGATGACATTACGATCGGTCAATCTAAGCAGCCT	60
QY	54	GlnPheLeuLysSerIleAspGluLeuAlaAlaPheSerValAlaValGluThrPheLys	73
Db	61	CAATTTTGAATCCATAGACGAATTAGCTCGGTTTTCAGTTGCAGTGGAAACATTCAA	120
QY	74	ArgGlnPheAspAspLeuGlnLysHisIleGluSerIleGluAsnAlaIleAspSerLys	93
Db	121	CGCCAAATTCGATGATGATTTCAGAGCACATCGAGTCAATCGAAACGCAATTGATTCAAA	180
QY	94	LeuGluSerAsnGlyValValLeuAlaAlaArgAsnAsnAsnPheHisGlnProMetLeu	113
Db	181	CTCAGAGTAGACGGGTGTCTCGCCGCGCGGACATATATTTCCATCAGCCGATGTTA	240
QY	114	SerProProArgAsnAsnValSerValGluThrThrValThrValSerGlnProSerGln	133
Db	241	TCGCCCTCCCGGAAACAATGTATCTGTAGAAACCAACCGTCACTGTGAGCCACCGTCTCAG	300
QY	134	GluIleValProGluThrSerAsnLysProGluGlyArgMetCysGluLeuMetCys	153
Db	301	GAGATTGTACCGGAGACGTCCGANTAAACCGAGGGGGAGGTATGTGTGATTGATGTGT	360
QY	154	SerLysGlyLeuArgLysTyrIleTyrAlaAsnIleSerAspGlnAlaLysLeuMetGlu	173
Db	361	AGCAAAGGTCTCGGTAAATACATATACGCGAATATCTCTGATCAAGCTTAAGTTAATGGAA	420
QY	174	GluIleProSerAlaLeuLysLeuAlaLysGluProAlaLysPheValLeuAspCysIle	193
Db	421	GAGATTCTTTCAGCTTTCGAATTCGCAAGGAGCGCAGGAGTTTGATTGGATTGATT	480
QY	194	GlyLysPheTyrLeuGlnGlyArgAlaPheThrLysGluSerProMetSerSerAla	213
Db	481	GGCAAGTTTACTTACAAAGGCGTAGAGCATTTACTTAAGAGTCGCTATGAGCTCTGCG	540
QY	214	ArgGlnValSerLeuLeu	219
Db	541	AGACAAAGTTTCGCTCTCT	558

Search completed: August 25, 2003, 12:30:04
Job time : 6688 secs

```

QY      174 GluileProSerAlaLeuLysLeuAlaLysGluProAlaLysPheValLeuAspCysIle 193
Db      421 GAGATTCCCTTCAGCTTTGAAATTGGCCAGGAGCCAGCAAGTTTGTATTGGATTGTATT 480
QY      194 GlyLysPheTyLeuGlnGlyArgAlaPheThrLysGluSerProMetSerSerAla 213
Db      481 GGCNAGTTTACTTACAGGGGTAGACATTTACTTAAAGAGTGCCTATGAGCTTCGG 540
QY      214 ArgGlnValSerLeuLeu 219
Db      541 AGACAAAGTTTCGCTCTT 558

RESULT 14
AY092691
LOCUS   AY092691
DEFINITION   Arabidopsis thaliana cultivar TSU-0 FRIGIDA protein gene, partial
cvs.
ACCESSION   AY092691
VERSION     AY092691.1
KEYWORDS    Arabidopsis thaliana (thale cress)
SOURCE      Arabidopsis thaliana
ORGANISM    Arabidopsis thaliana
REFERENCE   1 (bases 1 to 558)
AUTHORS    Hagenblad,J. and Nordborg,M.
TITLE      Sequence Variation and Haplotype Structure Surrounding the
JOURNAL    Flowering Time Locus FRI in Arabidopsis thaliana
MEDLINE    Genetics 161 (1), 289-298 (2002)
PUBMED     12019242
REFERENCE   2 (bases 1 to 558)
AUTHORS    Hagenblad,J. and Nordborg,M.
TITLE      Direct Submission
JOURNAL    Submitted (27-MAR-2002) Department of Genetics, Lund University,
Solvegatan 29, Lund SE-223 62, Sweden
FEATURES   source
            1..558
            /organism="Arabidopsis thaliana"
            /mol_type="genomic DNA"
            /cultivar="TSU-0"
            /db_xref="taxon:3702"
            <1..>558
            /product="FRIGIDA protein"
            /codon_start=1
            /protein_id="AA12144.1"
            /db_xref="GI:20160137"
            /translation="KIVETESTMDITIGQSKQFQFLKSIDELAAFSVAVETFKRQFD
            DIQKHIEIENAIKLSNGVLAARNNHQMLSPNNVSVETITVTSQPSQEI
            VPETSKNGGRMCMLCKSGKRYIYIANISDAKLMBEIPFSAKLAKAPEKFLVDCI
            GRFYLQGRATKESPMSSARQVSLI"
BASE COUNT 166 a 115 c 133 g 144 t
ORIGIN
Alignment Scores:
Pred. No.: 6,5e-53 Length: 558
Score: 931.00 Matches: 185
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.46% Mismatches: 0
Query Match: 29.99% Indels: 0
DB: 8 Gaps: 0

US-09-890-475-1 (1-609) x AY092691 (1-558)

QY      34 LysIleValGluThrGluSerThrSerMetAspIleThrIleGlyGlnSerLysGlnPro 53
Db      1 AAGATTGTGGAACAGAGTCTACAGTAGTGACATTACGATCGGTCAATCTAGCACGCT 60
QY      54 GlnPheLeuLysSerIleAspGlnLeuAlaLapheSerValAlaValGluThrPheLys 73

```

```

Db      61 CAATTTTGAATCCATAGACGAATTAGCTGCGTTTTCAGTTCAGTGGAAACATTCAAA 120
QY      74 ArgGlnPheAspLeuGlnLysHisIleGluSerIleGluAsnAlaIleAspSerLys 93
Db      121 CGCCAATTTCGATATATTTCAGAGCACATCGAGTCAATCGAAACGCAATTGATTCAAA 180
QY      94 LeuGluSerAsnGlyValValLeuAlaAlaArgAsnAsnAsnPheHisGlnProMetLeu 113
Db      181 CTCGAGAGTAACGGCGTTGCTCTCCGCGCGGCAACAATAATTTCATCAGCCGATGTTA 240
QY      114 SerProProArgAsnValSerValGluThrThrValThrValSerGlnProSerGln 133
Db      241 TCGCCTCCGCGGACATGATCTGTAGAACACACCGTCACTGTGAGCCACCGTCTCAG 300
QY      134 GluIleValProGluThrSerAsnLysProGluGlyGlyArgMetCysGluLeuMetCys 153
Db      301 GAGATTGTACCGGAGACGTCGAATAAAACCGGAGGGGAGCTATGTGTGAGTTGATGTGT 360
QY      154 SerLysGlyLeuArgLysTyIleTyAlaAsnIleSerAspGlnAlaLysLeuMetGlu 173
Db      361 AGCAAGGTCCTGGTAATATATATATACGGAATATCTCTGATCAAGCTAAGTTAATGGAA 420
QY      174 GluileProSerAlaLeuLysLeuAlaLysGluProAlaLysPheValLeuAspCysIle 193
Db      421 GAGATTCCCTTCAGCTTTGAAATTGGCCAGGAGCCAGCAAGTTTGTATTGGATTGTATT 480
QY      194 GlyLysPheTyLeuGlnGlyArgAlaPheThrLysGluSerProMetSerSerAla 213
Db      481 GGCNAGTTTACTTACAGGGGTAGACATTTACTTAAAGAGTGCCTATGAGCTTCGG 540
QY      214 ArgGlnValSerLeuLeu 219
Db      541 AGACAAAGTTTCGCTCTT 558

RESULT 15
AY092693
LOCUS   AY092693
DEFINITION   Arabidopsis thaliana cultivar Ler FRIGIDA protein gene, partial
cvs.
ACCESSION   AY092693
VERSION     AY092693.1
KEYWORDS    Arabidopsis thaliana (thale cress)
SOURCE      Arabidopsis thaliana
ORGANISM    Arabidopsis thaliana
REFERENCE   1 (bases 1 to 558)
AUTHORS    Hagenblad,J. and Nordborg,M.
TITLE      Sequence Variation and Haplotype Structure Surrounding the
JOURNAL    Flowering Time Locus FRI in Arabidopsis thaliana
MEDLINE    Genetics 161 (1), 289-298 (2002)
PUBMED     12019242
REFERENCE   2 (bases 1 to 558)
AUTHORS    Hagenblad,J. and Nordborg,M.
TITLE      Direct Submission
JOURNAL    Submitted (27-MAR-2002) Department of Genetics, Lund University,
Solvegatan 29, Lund SE-223 62, Sweden
FEATURES   source
            1..558
            /organism="Arabidopsis thaliana"
            /mol_type="genomic DNA"
            /cultivar="Ler"
            /db_xref="taxon:3702"
            <1..>558
            /product="FRIGIDA protein"
            /codon_start=1
            /protein_id="AA12146.1"
            /db_xref="GI:20160141"
            /translation="KIVETESTMDITIGQSKQFQFLKSIDELAAFSVAVETFKRQFD

```

CDS
/product="FRIGIDA protein"
<1..>558
/codon_start=1
/product="FRIGIDA protein"
/protein_id="AAW12134.1"
/db_xref="GI:20160117"

DIQHIENAIIDSKLENGVLAARNNNHOPMLSPNNVSVETTVVSQSQEI
/translation="KIVETESTMDITIGSQKQFOLKSIDELAASFVAVETFKRQFD
VPTSNKPEGRMCELMCKGLRKYIVANISDQAKLMEEIPSAKLAKPEAKFVLDICI
GKPYLQGRRAFTKESPMSSAROVSL"

BASE COUNT 166 a 115 c 133 g 144 t
ORIGIN

Alignment Scores:
Pred. No.: 6.5e-53 Length: 558
Score: 931.00 Matches: 185
Percent Similarity: 100.00% Conservativeness: 1
Best Local Similarity: 99.46% Mismatches: 0
Query Match: 29.99% Indels: 0
DB: 8 Gaps: 0

US-09-890-475-1 (1-609) x AY092681 (1-558)

QY 34 LysileValGluThrGluSerThrSerMetAspIleThrIleGlnSerLysGlnPro 53
Db 1 AAGATTGTCGAAACAGAGCTACAGTATGACATTCAGTACGTCGATCAATCAAGCAGCT 60
QY 54 GlnPheLeuLysSerIleAspGluLeuAlaPheSerValAlaValGluThrPheLys 73
Db 61 CAATTTTGAATCCATAGAGAAATAGTCGCTTTTTCAGTTCAGTGGAAACATTCAAA 120
QY 74 ArgGlnPheAspAspLeuGlnLysHisIleGluSerIleGluAsnAlaIleAspSerLys 93
Db 121 CGCCAAATTCGATATTCAGAACACATCGATCAATCGAAACGCAATTCATCCAAA 180
QY 94 LeuGluSerAsnGlyValValLeuAlaAlaArgAsnAsnAsnPheHisGlnProMetLeu 113
Db 181 CTCGAGAGTAAACCGGAGAGCTGCAATAAACCAGGAGGAGGAGCTATGCTGAGTGAATGCT 360
QY 114 SerProProArgAsnValSerValGluThrValThrValSerGlnProSerGln 133
Db 241 TCGCCTCCGCGGAGCAATGATCTGTGAACACCACTGAGTACGACCAACCGCTCTCAG 300
QY 134 GluileValProGluThrSerAsnLysProGluLysArgMetCysGluLeuMetCys 153
Db 301 GAGATTGTCGAGAGAGCTGCAATAAACCAGGAGGAGGAGCTATGCTGAGTGAATGCT 360
QY 154 SerLysGlyLeuArgLysTyrlleTyrlleTyrlleAlaAsnIleSerAspGlnAlaLysLeuMetGlu 173
Db 361 AGCAAGGTCGCGTAAATACATATACGCAATATCTGATCAAGCTAAGTTAATGGAA 420
QY 174 GluileProSerAlaLeuLysLeuAlaLysGluProAlaLysPheValLeuAspCysIle 193
Db 421 GAGATTCTTCAGCTTGAATTCGCCAGGAGGAGGAGCTATGCTGAGTGAATGCT 480
QY 194 GlyLysPheTyrlleGlnGlyArgArgAlaPheThrLysGluSerProMetSerSerAla 213
Db 481 GCGAAGTTTACTTACAGGCGTAGACATTTACTAAGAGTCGCTATGAGCTCTGCG 540
QY 214 ArgGlnValSerLeuLeu 219
Db 541 AGCAAGTTTCGCTTCTT 558

RESULT 13
LOCUS AY092683
DEFINITION Arabidopsis thaliana cultivar GOT-32 FRIGIDA protein gene, partial cds.
ACCESSION AY092683
VERSION AY092683.1 GI:20160120
KEYWORDS Arabidopsis thaliana (thale cress)
SOURCE Arabidopsis thaliana
ORGANISM Arabidopsis thaliana

REFERENCE
AUTHORS Hagenblad, J. and Nordborg, M.
TITLE Sequence Variation and Haplotype Structure Surrounding the Flowering Time Locus FRI in Arabidopsis thaliana
JOURNAL Genetics 161 (1), 289-298 (2002)
MEDLINE 22013871
PUBMED 12019242

REFERENCE
AUTHORS Hagenblad, J. and Nordborg, M.
TITLE Direct Submission
JOURNAL Solvegatan 29, Lund SE-223 62, Sweden

FEATURES
source
1..558
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/cultivar="GOT-32"
/db_xref="taxon:3702"
<1..>558
/product="FRIGIDA protein"
<1..>558
/codon_start=1
/product="FRIGIDA protein"
/protein_id="AAW12134.1"
/db_xref="GI:20160117"

BASE COUNT 166 a 115 c 133 g 144 t
ORIGIN

Alignment Scores:
Pred. No.: 6.5e-53 Length: 558
Score: 931.00 Matches: 185
Percent Similarity: 100.00% Conservativeness: 1
Best Local Similarity: 99.46% Mismatches: 0
Query Match: 29.99% Indels: 0
DB: 8 Gaps: 0

US-09-890-475-1 (1-609) x AY092683 (1-558)

QY 34 LysileValGluThrGluSerThrSerMetAspIleThrIleGlnSerLysGlnPro 53
Db 1 AAGATTGTCGAAACAGAGCTACAGTATGACATTCAGTACGTCGATCAATCAAGCAGCT 60
QY 54 GlnPheLeuLysSerIleAspGluLeuAlaAlaPheSerValAlaValGluThrPheLys 73
Db 61 CAATTTTGAATCCATAGAGAAATAGTCGCTTTTTCAGTTCAGTGGAAACATTCAAA 120
QY 74 ArgGlnPheAspAspLeuGlnLysHisIleGluSerIleGluAsnAlaIleAspSerLys 93
Db 121 CGCCAAATTCGATATTCAGAACACATCGATCAATCGAAACGCAATTCATTCAAA 180
QY 94 LeuGluSerAsnGlyValValLeuAlaAlaArgAsnAsnAsnPheHisGlnProMetLeu 113
Db 181 CTCGAGAGTAAACCGGAGTTCCTCGCGGAGGAGGAGCTATGCTGAGTGAATGCT 240
QY 114 SerProProArgAsnValSerValGluThrValThrValSerGlnProSerGln 133
Db 241 TCGCCTCCGCGGAGCAATGATCTGTAGAAACACCGTCACTGTGAGCCACCGTCTCAG 300
QY 134 GluileValProGluThrSerAsnLysProGluLysArgMetCysGluLeuMetCys 153
Db 301 GAGATTGTCGAGAGAGCTGCAATAAACCAGGAGGAGGAGCTATGCTGAGTGAATGCT 360
QY 154 SerLysGlyLeuArgLysTyrlleTyrlleTyrlleAlaAsnIleSerAspGlnAlaLysLeuMetGlu 173
Db 361 AGCAAGGTCGCGTAAATACATATACGCAATATCTGATCAAGCTAAGTTAATGGAA 420

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

Db 76313 CGTAGAGCAATTTACTAAAGAGTCGCTATGAGCTCTGCGAGACAAGTTTCGCTCTTATA 76372
 Qy 221 LeuGluSerPheLeuLeuMetProAspArgGlyLysGlyValLysValLysLeuSerTip 240
 Db 76373 CTGGAGCTCTTTCTCTTAATGCTGATCGTGTAAAGGGAAGGTGAAGATTGAGAGTTGG 76432
 Qy 241 IleLysAspGluAlaGluThrAlaAlaValAlaTTPArgLysArgLeuMetThrGluGly 260
 Db 76433 ATTAAAGATTGAGCGGAGACGGCTGCTGTGCTTGGAGGAAAGGTGATGACTGAAGGA 76492
 Qy 261 GlyLeuAlaAlaAlaGluLysMetAspAlaArgGlyLeuLeuLeuValAlaCysPhe 280
 Db 76493 GGATTAGCTGCGCTGAGAAAATGATGCAAGGGGTTCCTTTACTAGTTGCTTTT 76552
 Qy 281 GlyValProSerAsnPheArgSerThrAspLeuLeuAspLeuLeuArgMetSerGlySer 300
 Db 76553 GGTGTTTCCTTCAAACTTTAGGAGTACAGATTTCGTGGATTTGATAAGGATGAGTGTTCG 76612
 Qy 301 AsnGluIleAlaGlyAlaLeuLysArgSerGlnPheLeuValProMet----- 316
 Db 76613 AATGAGATTGCCGGTGCCTTGAAGCGGTACAGTTTCAGTTACCATATTTCTGTCTCTCAC 76671
 Qy 316 ----- 316
 Db 76672 TCGGTCAATTTCAATGCCAAAGTGGTCTCTTTTGTGATCATCGACCAACATCAAGTT 76731
 Qy 316 ----- 316
 Db 76732 CCATCTTTGTTTTTCGATAAGCTTGATGGTATAACTAGGAGAGACATCAATATTTAG 76791
 Qy 316 ----- 316
 Db 76792 AGTGAATGACTGATTGAGCCAAATCCTAGCTAGAAATTAATCTGGAAGAACTTGAAC 76851
 Qy 316 ----- 316
 Db 76852 TCTCAACCATAGTTTGGTACGAAATTTGTTGCTGTCAGAACCAAAATGATAGGCTATTG 76911
 Qy 316 ----- 316
 Db 76912 CCTTGAATAGTGTCTTGTGTTTCCAAATTATGGAAGTTAAATCATATGACTTAGCT 76971
 Qy 316 ----- 316
 Db 76972 GTTGATCTAATTAAGCTTAAGCAATGCCAACTCTAAGAGTGGTACTTACACAATATT 77031
 Qy 317 -----ValSerGlyValGluSerSerIleLysArgGlyMetHisIleGluAlaLeu 334
 Db 77032 CTATTGGTCATAGGTATAGTTGAATCAAGTATCAAGCGTGAATGCATATTGAAGCTCTT 77091
 Qy 335 GluMetValTyThrPheGlyMetGluAspLysPheSerAlaLeuValLeuThrSer 354
 Db 77092 GAAATGGTTTATACCTTTGGCATGGAGATAAGTTTTCAGTGTCTAGTTCTAACTTCA 77151
 Qy 355 PheLeuLysMetSerLysGluSerPheGluArgAlaLysArgLysAlaGlnSerProLeu 374
 Db 77152 TTTCTTAAGATGAGCAAGAGTCAATTGAGAGCGCAAAACGGAAGCCCAAGTCAACGCTG 77211
 Qy 375 AlaPheLys----- 377
 Db 77212 GCATTTGT-ATGAACCTTCCCTTGACATTATGACTTTATGAACCTTTTATCATCAT 77270
 Qy 378 -----GluAlaAlaThrLysGlnLeu 384
 Db 77271 CTGAGCTGACCATGATATATTATTCTCAACAGAAAGCGGTCAAAAGCAGCTA 77330
 Qy 385 AlaValLeuSerSerValMetGlnCysMetGluThrHisLysLeuAspProAlaLysGlu 404
 Db 77331 GCTGTGTTATCATCAGTTATGCAAGTATGAGACTCACAAGTTAGATCTCTGCGAAGAA 77390
 Qy 405 LeuProGlyTyThrGlnIleLysGluGlnIleValSerLeuGluLysAspThrLeuGlnLeu 424
 Db 77391 CTACCAGGATGCGAGATCAAGAGCAAAATTTGTTAGCTTGGAGAAAGACATCTTTCAGCTC 77450

Qy 425 AspLysGluMetGluGluLysAlaArgSerLeuSerLeuMetGluGluAlaLeuAla 444
 Db 77451 GACAAAGAGATGGAAGAGACAAGCAAGATCTCTCAGTTTAAATGGAGGAAGCGCACATTGCC 77510
 Qy 445 LysArgMetTyAsnGlnGlnIleLysArgProArgLeuSerProMetGluMetProPro 464
 Db 77511 AAGAAGAAATGTATAACCAACAGATAAAACGTCACAGGTTATCACCCATGGAATGCCACCA 77570
 Qy 465 ValThrSerSerSerTySerProIleTyArgAspArgSerPheProSerGlnArgAsp 484
 Db 77571 GTAACCTTCTTCATCGTATTCCTTATCTACCGTGATAGAAGCTTTCCTAGTCAAGAGAC 77630
 Qy 485 AspAspGlnAspGluIleSerAlaLeuValSerSerTyLeuGlyProSerThrSerPhe 504
 Db 77631 GATGACCAAGATGAATATATCATCAGCTCTTGTGAGTAGTTACCTCGCGCCGTCACATCTTTT 77690
 Qy 505 ProHisArgSerArgAspSerProGluTyMetValProLeuProHisGlyGlyLeuGly 524
 Db 77691 CTTCTATCCTCAAGAGAGATCCCGGAATATATGTTTCCATTTCCACATGTTGGTTAGGA 77750
 Qy 525 ArgSerValTyAlaTyArgGluHisLeuAlaProAsnSerTySerProGlyHisGlyHis 544
 Db 77751 AGAGTGATATATGATATGAACATCTGCGCCCAAAATCATATTTCTCAGGTCAAGGACAT 77810
 Qy 545 ArgLeuHisArgGlnTySerProSerLeuValHisGlyGlnArgHisProLeuGlnTy 564
 Db 77811 AGACTTATCATCAGAGTACTCTCCGCTCTTGTGTTACGAGAGAGATCCACTACAGTAC 77870
 Qy 565 SerProPheHisGlyGlnGlnLeuProTyArgIleGlnArgValTyArgHis 584
 Db 77871 TCTCTCCAAATTCATGACAAACACAGTTACCATATGGTATACAAAGGTTTACACACAT 77930
 Qy 585 SerProSerGluGluArgTyLeuGlyLeuSerAsnGlnArgSerProArgSerAsnSer 604
 Db 77931 TCACCATCTGAAGAAAGATATTGGGTTTATCCAAATCAAAGGTTCTCTCGCAGTAACTCA 77990
 Qy 605 SerLeuAspProLys 609
 Db 77991 TCATTAGACCCCAAA 78005
 RESULT 12
 LOCUS AY092681
 DEFINITION Arabidopsis thaliana cultivar RSCH-4 FRIGIDA protein gene, partial cds.
 ACCESSION AY092681
 VERSION AY092681.1 GI:20160116
 KEYWORDS Arabidopsis thaliana (thale cress)
 SOURCE Arabidopsis thaliana
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 REFERENCE 1 (bases 1 to 558)
 AUTHORS Hagenblad, J. and Nordborg, M.
 TITLE Sequence Variation and Haplotype Structure Surrounding the Flowering Time Locus FRI in Arabidopsis thaliana
 JOURNAL Genetics 161 (1), 289-298 (2002)
 MEDLINE 12019242
 PUBMED 12019242
 REFERENCE 2 (bases 1 to 558)
 AUTHORS Hagenblad, J. and Nordborg, M.
 TITLE Direct Submission
 JOURNAL Submitted (27-NOV-2002) Department of Genetics, Lund University, Solvegatan 29, Lund SE-223 62, Sweden
 FEATURES
 source
 1. 558
 /organism="Arabidopsis thaliana"
 /molecule="genomic DNA"
 /cultivar="RSCH-4"
 /db_xref="taxon:3702"
 <1..558
 mRNA

```

intron 11019..11078
/gene="AT4g00450"
/number=4
exon 11079..11423
/gene="AT4g00450"
/number=5
intron 11424..11474
/gene="AT4g00450"
/number=5
exon 11475..13674
/gene="AT4g00450"
/number=6
intron 13675..13734
/gene="AT4g00450"
/number=6
exon 13735..13823
/gene="AT4g00450"
/number=7
intron 13824..13920
/gene="AT4g00450"
/number=7
exon 13921..16530
/gene="AT4g00450"
/number=8
intron 16531..16676
/gene="AT4g00450"
/number=8
exon 16677..16862
/gene="AT4g00450"
/number=9
intron 16863..16975
/gene="AT4g00450"
/number=9
exon 16976..17042
/gene="AT4g00450"
/number=10
intron 17043..17091
/gene="AT4g00450"
/number=10
exon 17092..17547
/gene="AT4g00450"
/number=11
gene 18027..19004
/gene="AT4g00460"
complement(join(18027..18146,18234..18578,18669..19004))
/gene="AT4g00460"
complement(join(18027..18146,18234..18578,18669..19004))
/gene="AT4g00460"
contains EST gb:AI993831.1
/codon_start=1
/product="putative protein"
/protein_id="CAB80855.1"
/db_xref="GI:7267385"
/translation="MWTISFDPCLLQELADSDSFONTEFWAEEGSLSMKARSSTGSFR
KVIQREBKWLFPVLPVSEGLSDKARKQLKNRSTNQIHKAAVAINSSILSMEI
PDSYMTLPKCGSSVSDSIYRYMSGGRFFPEQLDCLNI SSEHVAQLADRVASM
YTWRRKSLNSKNSNMWVMDLMTSTTRTDKNVMAERAETLLFCLKQRYPELSQTSI
DICKIQNKDVGKAVLESYSRVLEGLAFNIVAMIDDLVLYDKTMRGSE"
complement(18027..18146)
/gene="AT4g00460"
/number=1
intron complement(18147..18233)
/gene="AT4g00460"
/number=1
exon complement(18234..18578)
/gene="AT4g00460"
/number=2
intron complement(18579..18668)
/gene="AT4g00460"
/number=2
exon complement(18669..19004)
/gene="AT4g00460"

/number=3
19070..19781
/gene="AT4g00470"
complement(join(19070..19221,19270..19386,19472..19781))
/gene="AT4g00470"
complement(join(19070..19221,19270..19386,19472..19781))
/gene="AT4g00470"
/notes="similarity to
contains EST gb:AA721856"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAB80856.1"
/db_xref="GI:7267386"
/translation="MENLSNFDENDHQSRSIDQNSQSAVETEVYTMISIDSEVYPR
TCSETTSPGSDQIDETNSFCSEASPCNWPVLTEGSKSKCLSGLEMQNECLVGEISE
PELEWTKERFAKLLGEDMSGSGKGVCTAVTISNAITLNYDNCNKDTSFDNVCSYSVM
TSEVRAVRNRESIVERNELSFICMLHCR1"
complement(19070..19221)
/gene="AT4g00470"
/number=1

exon

Alignment Scores:
Pred. No.: 1,08e-181 Length: 197975
Score: 2936.00 Matches: 603
Percent Similarity: 78.95% Conservative: 1
Best Local Similarity: 78.82% Mismatches: 5
Query Match: 94.59% Indels: 158
DB: 8 Gaps: 2

US-09-890-475-1 (1-609) x ATCHRV2 (1-197975)

Qy 1 MetSerAsnTy:ProProThrValAlaGlnProThrThrThrAlaAsnProLeuLeu 20
Db 75713 ATGTCCAATTATCCACCGCGGTGGCGCAACCAACAGACGCGAATCCACTGCTG 75772
Qy 21 GlnArgHisGluSerGluGlnArgArgGluLeuProLysIleValGluThrGluSer 40
Db 75773 CAGGACATCAATCTGAACAGCGACGAGAGATTACCGAAGATTGTGGAACAGAGTCT 75832
Qy 41 ThrSerMetAspIleThrIleGlyGlnSerLysGlnProGlnPheLeuLysSerIleasp 60
Db 75833 ACAAGTATGGACATTACGATCGGTCAATTAAGACGCGCTCAATTTTGAATCCATAGAC 75892
Qy 61 GluLeuAlaAlaPheSerValAlaValGluThrPheLysArgGlnPheAspAspLeuGln 80
Db 75893 GAATTAGCTCGCTTTTCAGTTCGAGTGGAAACATTAACACGCCAATTCGATGATCTTCAG 75952
Qy 81 LysHisIleGluSerIleGluAsnAlaIleAspSerLysLeuGluSerAsnGlyValVal 100
Db 75953 AAGCACATCGAGTCAATCGAAAACGCAATTGATTCCAAACTCGAGAGTAACGGCGTTGTC 76012
Qy 101 LeuAlaAlaArgAsnAsnAsnPheHisGlnProMetLeuSerProArgAsnAsnVal 120
Db 76013 CTGCGCGCGGAGAACAAATTAATTCATAGCCGATGTTATCGCTCCCGGGAACATGTA 76072
Qy 121 SerValGluThrThrValThrValSerGlnProSerGlnGluIleValProGluThrSer 140
Db 76073 TCTGTAGAACCACCGCTACTGTGAGCAACCGCTCTCAGGAGATTGTACCAGAGACGCTG 76132
Qy 141 AsnLysProGluGlyGlyArgMetCysGluLeuMetCysSerLysGlyLeuArgLysTyr 160
Db 76133 AATAAACCGGAGGGGAAACGTATGTAGTTGATGTGTAGCAAAAGGTTCTCGTAATAATAC 76192
Qy 161 IleTyrAlaAsnIleSerAspGlnAlaLysLeuMetGluGluIleProSerAlaLeuLys 180
Db 76193 ATATACGCAATATCTCTGATCAAGCTAAGTTAATGGAAGAGATTCTTCAGCTTTGAA 76252
Qy 181 LeuAlaLysGluProAlaLysPheValLeuAspCysIleGlyLysPheTyrIleuGlnGly 200
Db 76253 TTGCCAAGGAGGACCGAAGTTTGTATTGGATTGTATTGGCAAGTTTACTTCAAGGG 76312
Qy 201 ArgArgAlaPheThrLysGluSerProMetSerSerAlaArgGlnValSerLeuLeuIle 220

```

```

QY 445 LysArgMetTyrAsnGlnGlnIleLysArgProArgLeuSerProMetGluMetProPro 464
D 58637 AAGAGATGATTAACCAACAGATTAACAGTCAAGGTATATCCCATCGAATGCCACCA 58696
QY 465 ValThrSerSerTyrSerProIleTyrArgAspArgSerPheProSerGlnArgAsp 484
D 58697 GTAACCTCTTCATCGTATCTCTATCTACCGTATAGAGAGCTTTCTCTAGTCAAAGAGAC 58756
QY 485 AspAspGlnAspGlnLeuSerSerTyrLeuGlyProSerThrSerPhe 504
D 58757 GATGACCAAGATGAATATACAGCTCTTGAGAGTAGTACCTCGGCCGCTCAACATCTTT 58816
QY 505 ProHisArgSerArgSerProGlnTyrMetValProLeuProHisGlyGlyLeuGly 524
D 58817 CCTCATCGCTCAAGAAGATCCCGGAATATATGTTCCATCTCCATGTTGGTGTAGGA 58876
QY 525 ArgSerValTyrAlaTyrGluHisLeuAlaProAsnSerTyrSerProGlyHisGlyHis 544
D 58877 AGAAGTGATATGATATGATATGATATGATATGATATGATATGATATGATATGATATGAT 58936
QY 545 ArgLeuHisArgGlnTyrSerProSerLeuValHisGlyGlnArgHisProLeuGlnTyr 564
D 58937 AGACTTCATCGACAGTACTCTCGGCTTTGGTTACGGACAGACATCCACTACAGTAC 58996
QY 565 SerProProIleHisGlyGlnGlnGlnLeuProTyrGlyIleGlnArgValTyrArgHis 584
D 58997 TCTCCCTCAATTCATGACATGACATGACATGACATGACATGACATGACATGACATGAC 59056
QY 585 SerProSerGluGluArgTyrLeuGlyLeuSerAsnGlnArgSerProArgSerAsnSer 604
D 59057 TCACCATCTGAGAAGATATTTGGGTTATTCATCAATCAAGGCTCTCTCGCAGTAATCA 59116
QY 605 SerLeuAspProLys 609
D 59117 TCATTAGACCCCAAA 59131

RESULT 11
LOCUS ATCHRIV2 197975 bp DNA linear PLN 16-MAR-2000
DEFINITION Arabidopsis thaliana DNA chromosome 4, contig fragment No. 2.
ACCESSION AL161472
VERSION AL161472.2 GI:7267383
KEYWORDS
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 197975)
AUTHORS Lamar, B., Stoneking, T., Stumpf, J., Mewes, H.W., Lemcke, K. and
Mayer, K.F.X.
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 197975)
AUTHORS EU Arabidopsis sequencing project.
TITLE Direct Submission
JOURNAL Submitted (10-MAR-2000) MIPS, at the Max-Planck-Institut fuer
Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail:
lemcke@mips.biochem.mpg.de, mayer@mips.biochem.mpg.de
Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge
Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,
E-mail: michael.bevan@bbsrc.ac.uk
COMMENT Information on performance of analysis and a more detailed
annotation of this entry and other sequences of chromosomes 3, 4
and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/
this fragment has an overlap with ATCHRIV1 at the 5' end and an
overlap with ATCHRIV3 at the 3' end.
FEATURES
source
1..197975
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/variety="Columbia"
/db_xref="taxon:3702"

```

```

/chromosome="4"
10352..17547
/gene="AT4g00450"
Join(10352..10384,10432..10545,10605..10709,10849..11018,
11079..11423,11475..13674,13735..13823,13921..16530,
16677..16862,16976..17042,17092..17547)
/gene="AT4g00450"
/note="Contains aminoacyl-transfer RNA synthetases
class-II signatures AA1343-1352; Prokaryotic membrane
lipoprotein lipid attachment site AA349-359
contains EST gb:AI999580.1"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAB80854.1"
/db_xref="GI:7267384"
/translators="MSPLEFVISTGAVKCLRAINESRALKRVKIKELSDSNPQRTIF
CAEPRAGVIGVPLSGSLLCAPFFPEQSGEETKRWIESLSCQHKRLKSLADNPFG
YRRKTLFVIRNNVPLLRATWFKVYTLNQVHCWAINCDLVLFLQKHLTRHKL
GVSGQKMLNICTNWNFCGIAHLLSKLEIGHRCFIQCKRGIQVQQAFTAR
KELYLINGIWCVFSGTLMKGFLLISLITGFSSYSFVNFHALQEKEIFEIQL
LLPIVYGESIVLSQTVQSLVAIVRFQEPAPGSGDLVDSNRATYLSAIEWR
YLVAAPDTVASDPFLPPSVAACPNDSYTSKAVENLEKLSNABEISAOFGRR
VLSREFTSFDTYTTICRSADDLAKIASAGYQCHNVAKVQALDKALSGDIFRAVS
VLFEDLNGAVDEAWITDVSPCLASSLRWGAISTSVFVFELEWATCDFRDFRAG
VKOTIKSGRDCSQVILVLIQLLKQILGGEFTARKGNKNRNNFLGVKSGSMDAPE
SPGLHDIIIVMDIOHEVHGKAKRLQLVFLIRSGIFNPIAYVQLIIVSGMIDVIQ
PAVDFERRMRHRIKQLPGCFVHETLEEAQLFGGDKLSEAVRTYSNERLLRELLV
EKGYNNKLLSDQKSIKSTSSVFPACNAKSGKPRKTKSVDIRLEKRI
SALLQYPMQSGVETPVDRFONSVKRSGSVYSKMDPEATPCEDCRCKRKNND
EKSSCYQNSPIASDEEDNMWIKGSKTVESLKVDPQIETIKOVRGRCKMARKTQS
LAQCAARLESGASTSHVCDNKVSPHGGPVEGENQKVDVFTSTFVMSVGN
SLKQLQVKKSIATVLTITVRQLVSEFPQKSVRVGQFNAGPVEKNTIRWLGKADE
LYSLFLDLSDLVSAAGRNLTVPNRNENMCEGEALVSSLRYENILLISADLV
PEAMTALNRAASLMSNGKISGSAALVTRYILKRYGLSPVVEHNNFKATSEKL
LSELDRSGNGEYGNLPGVAGVNDPDDYLRKISIGGARPSRVGLMREDVLQRHVE
EATHYLLKLIQGTWKASLAEKNDGQVAAQIVVGLMDCIROTGGAOBDPSLVSS
QVSAIINSGVLSVARIITDFSLGNIYONHPGSDSNRIARVLIHITCLLKEALGE
RQSVFEIALATESSTALITGVFAPVKSRQHQSLSPESYDSNANNSTIDMSNGTQMA
LSRAKTAIVASVAVIGITHTGVITLERIVGLRLKDYLDFOVFARKSSNGSRA
MGAKRSFPIVYVHVRFLVGNCKTVSEGLVLELVGESSVAISRMLPLKLVFP
PAYISITFVLRVFPVSNNSNSVHEDTHRYLSLTMAFDHVIKHPFRDVCFRDTQG
LYEITVADSTDAEFASVESHGDMHLKSVAFAPLARFLNSLIDCKVPSSGYHMEG
VSEAKRHQNGTKLVKLVSLDCLQAPKQWVLELLLEQLAEKLENDHML
TDAIRSCPTSEKPDASENEKNFTQILLTLVLRPDAVPLFSVHLFGESVDSMLK
QAEWFLAGQDVLFGKTIROKLIIVGESKGLTPKQPKWPGWNCNSSSSSHITANKAG
KCKRFEITSEEGEVEIEGSRKVLPRVIDENSPVGYGITERAFVOLLVLPDQ
SSDSRSTFVNLVRFQSNIEQLQSLVYNSRTTSNKQMTAGSSGSEISNKGSTRKL
RGSPSLARRSSANTDTPPPPALRASMLRQLFLRLPVIICEPFSFKTRHAL
ASTVRLIGSRVYEDYVAGSPSELGSKAETESTIDPSSMADLSSEVLPFRRLTGLS
TNFKSLKVLGCKNELSRMOLPDTIRWRIQAAMPILLPLSLRSLSCOPHSVPPTAL
TLVOPSGSTAAAGTNQNSPAISKSGTAAAGKLPKTMPLAPHQOQEQADNTDVPDPTL
LEDGTSGLSSSNASNSDANLRAATCWLKGVAVRVRTDITYVGSVDDDS"
10352..10384
/gene="AT4g00450"
/number=1
10385..10431
/gene="AT4g00450"
/number=1
10432..10545
/gene="AT4g00450"
/number=2
10546..10604
/gene="AT4g00450"
/number=2
10605..10709
/gene="AT4g00450"
/number=3
10710..10848
/gene="AT4g00450"
/number=3
10849..11018
/gene="AT4g00450"
/number=4

```

```

/ gene="F6N23.19"
Complement(join(22410..22521,23220..23538,23620..23939,
24009..24137,24240..24352,24439..24548,24639..24753))
/ gene="F6N23.19"
/codon_start=1
/evidence=not experimental
/protein_id="AAC13637.1"
/db_xref="GI:3047126"
/t_rna_lations="MSPSYINVSLKPPFPHLMQDQILTNVVKYGRNWKRIAECLPG
SENRNDVOCQRHLKVLDPDLQKGMKEDELLSELVDKYMENDPPWKSISKEL
PGRIGQCRWRNHLNPTIISFWTREBELILVQQRNGNKNWAEIAKLPLPRTENN
IKHWCVKRLEQFNLGVVYKSPSSGFYFNFNQNTNWESCITTSQIKRKA
KSPQRLDILTLGLNWRSSSTSLRGEESVSDVCAELNACLTPQNSNNDTA
CWEVEMKERLMAARTFTPSIIISKTSPASGLKRLQKYDTPFPDARSHMSSE
DLXSVASPSKSYFVKRNTCSGKPLERLDFDLWDBHGRNGIVNFSVRILLPKS
DLXSLVRPFWLR"

Alignment Scores:
Pred. NO.: 3.93e-182 Length: 91849
Score: 2936.00 Matches: 603
Percent Similarity: 78.95% Conservative: 1
Best Local Similarity: 78.82% Mismatches: 5
Query Match: 94.59% Indels: 158
DB: 8 Gaps: 2

US-09-890-475-1 (1-609) x F6N23 (1-91849)

QY 1 MetSerAsnTyrProProThrValAlaAlaGlnProThrThrThrAlaAsnProLeuLeu 20
DB 56839 ATGTCCAAATTATCCACGCGGTGGCGGCAACCCACACGACGCGCAATCCACTGCTG 56898
QY 21 GlnArgHisGlnSerGluGlnArgArgGluLeuProLysIleValGluThrGluSer 40
DB 56899 CAGCGACATCAATCTGAACAGCAGCAGAGAGATTACCGAAGATTCTCGAACAGAGTCT 56958
QY 41 ThrSerMetAspIleThrIleGlyGlnSerLysGlnProGlnPheLeuLysSerIleAsp 60
DB 56959 ACAAGTATGGACATTAGATCGGTCAATCTAAGCAGCCTCAATTTTGAATCCATAGAC 57018
QY 61 GluLeuAlaAlaPheSerValAlaValGluThrPheLysArgGlnPheAspAspLeuGln 80
DB 57019 GAATTAGTGGTGTTCAGTTGAGTGGAAACATTCAACGCGCAATTCGATGATCTTCAG 57078
QY 81 LysHisIleGluSerIleGluAsnAlaIleAspSerLysLeuGluSerAsnGlyValVal 100
DB 57079 AAGCACATCGAGTCAATCGAAACGCAATTGATTCCAAACCTCAGAGATTAACGCGTGTGC 57138
QY 101 LeuAlaAlaArgAsnAsnAsnPheHisGlnProMetLeuSerProProArgAsnAsnVal 120
DB 57139 CTCGCGCGCGGAACAATAATTTCCATCAGCCGATGTTATCGCCTCCGCGGAACAATGTA 57198
QY 121 SerValGluThrThrValThrValSerGlnProSerGlnGluIleValProGluThrSer 140
DB 57199 TCTGTAGAACACCGTCACCTGTGAGCGCAACCGTCTCAGAGAGATTGTACCGAGAGCTGC 57258
QY 141 AsnLysProGluGlyArgMetCysGluLeuMetCysSerLysGlyLeuArgLysTyr 160
DB 57259 AATAAACCGGAGGGGAACGATATATGTGAGTTGATGTAGTGGTAAAGGCTCTGGCTAAATAC 57318
QY 161 IleTyrAlaAsnIleSerAspGlnAlaLysLeuMetGluIleProSerAlaLeuLys 180
DB 57319 ATATACGGGAATATCTCTGATCAAGCTAAGTTAATGAAGAGATTCTTCAGCTTTGAAA 57378
QY 181 LeuAlaLysGluProAlaLysPheValLeuAspCysIleGlyLysPheTyrLeuGlnGly 200
DB 57379 TTGGCCAGGAGGCGAGCGAAGTTTGTATTCGATTGTTATTCGCAAGTTTACTTACAAGG 57438
QY 201 ArgArgAlaPheThrLysGluSerProMetSerSerAlaArgGlnValSerLeuIle 220
DB 57439 CGTAGACATTATCTAAAGAGTCCGCTATGAGTCTCGGAGACAAGTTTCGCTCTTTATA 57498
QY 221 LeuGluSerPheLeuLeuMetProAspArgGlyLysGlyLysValLysIleGluSerTyr 240

```

```

DB 57499 CTGAGCTCTTTTCTTCTTAATGCTCTGATCGTGTAAAGGAAAGGTCAAGATTGAGATTGG 57558
QY 241 IleLysAspGluAlaGluThrAlaAlaValAlaIleArgLysArgLysLeuMetThrGluGly 260
DB 57559 ATTAAAGATGAGCGGAGACGGCTCTGTTCTTGGAGAAAAGGTGATGACTGAAGGA 57618
QY 261 GlyLeuAlaAlaGluLysMetAspAlaArgGlyLeuLeuLeuValAlaCysPhe 280
DB 57619 GGATTAGCTCGGCTGAGAAAATGATGCAAGGGGTTTGTCTTTACTAGTTGCTGTGTTT 57678
QY 281 GlyValProSerAsnPheArgSerThrAspLeuLeuAspLeuIleArgMetSerGlySer 300
DB 57679 GGTGTTCTTCTCAACCTTAGGAGTACAGATTCTGGATTGATGAAGATGAGTGTTCG 57738
QY 301 AsnGluIleAlaGlyAlaLeuLysArgSerGlnPheLeuValProMet 316
DB 57739 AATGAGATTGCCGGTCTTTGAAGCGGTACAGTTTCAGGTACCATTA-TTCTGTTCTCAC 57797
QY 316 ----- 316
DB 57798 TCGTGGAATTTCAATGCAAAAGGTGTTCTTTTGTGTGACATCATCGACCAACATCAAGTT 57857
QY 316 ----- 316
DB 57858 CCATCTTGTGTTTCGATAAGCTTGTATGCTATAAACTAGGAGACACATCAATATTATTAG 57917
QY 316 ----- 316
DB 57918 AGTGCAATGACTGATTGACCCAAATCTTAGCTAGAGAAATTAATCTGGAAAGAACTTGGAAC 57977
QY 316 ----- 316
DB 57978 TCTCAACCATAGTGTGTTGGTAGCAAAATGTTGCTTGTTCAGAACCAATGATAGCTATTG 58037
QY 316 ----- 316
DB 58038 CCTTGAATAGTGTGTTCTTGTGGTTCCTCAATATTGGAAGTTAAAAATCATATGACTTAGCT 58097
QY 316 ----- 316
DB 58098 GTTGGATACTAATTAGCTTAAGCAATGCCAATCTAAGAAAGTGGTACTTACACAATATT 58157
QY 317 -----ValSerGlyIleValGluSerSerIleLysArgGlyMetHisIleGluAlaLeu 334
DB 58158 CTATTGGTCATAGTATAGTTGAATCAAGTATCAAGCGTGAATGCATATTGGAAGCTCTT 58217
QY 335 GluMetValTyrThrPheGlyMetGluAspLysPheSerAlaAlaLeuValThrSer 354
DB 58218 GAAATGGTTTATACCTTTTGGCATGGAGGATAGTTTTCAGCTGCTCTAGTTCTTACTTCA 58277
QY 355 PheLeuLysMetSerLysGluSerPheGluArgAlaLysArgLysAlaGlnSerProLeu 374
DB 58278 TTCTTAAAGATGAGCAAGAGTCAATTGAGAGGGCAAAACGMAAGCCAGTACCCTG 58337
QY 375 AlaPheLys----- 377
DB 58338 GCATTTGT-ATGAACCCCTTCCTTCGACATATGTAACCTTTATGAACTCTTTATCATCAT 58396
QY 378 -----GluAlaAlaThrLysGlnLeu 384
DB 58397 CTGAGCTGACCATTCATATATTTTCTCAACAGAAAGAGCGCTACAAAGCAGCTA 58456
QY 385 AlaValLeuSerSerValMetGlnCysMetGluThrHisLysLeuAspProAlaLysGlu 404
DB 58457 GCTGTGTTATCATCATGATTATGAGTGTATGAGAGATCAAGATTAGTTCCTCGCAAGAA 58516
QY 405 LeuProGlyTyrGlnIleLysGluGlnIleValSerLeuGluLysAspThrLeuGlnLeu 424
DB 58517 CTACAGGATGGCAGATCAAGAGCAATTTGTAGCTTGGAGAAAGACACTCTTCAGTCT 58576
QY 425 AspLysGluMetGluGluLysAlaArgSerLeuSerLeuMetGluGluAlaAlaLeuAla 444
DB 58577 GACAAAGAGATGGAAGAAAGCAAGATCTCTCAGTTTAAATGGAGAAAGCCGACCTTCC 58636

```


by M. Marra, WashU, to pick the best candidates for sequencing.

NOTICE: This sequence may not be the entire insert of this clone. It may be shorter because we only sequence overlapping sections once, or longer because we provide a small overlap between neighboring submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone

NEFBORING COSMID INFORMATION:

The 5' clone is F5110, 200 bp overlap; 3' clone is F15P23, 900 bp overlap. Actual start of this clone is at base position 104960 of CELF5110; actual end is at 91149 of CELF6N23

The clone sequenced to the left is F5110. The actual start of this clone is at base position 104960 of F5110; actual end is at base position 91149 of F6N23.

NOTES:

Coding sequences below are predicted from computer analysis, using the program GeneFinder (P. Green and L. Hillier, ms in preparation).

FEATURES

```

source
    i. 31849
    /organism="Arabidopsis thaliana"
    /mol_type="genomic DNA"
    /cultivar="Columbia"
    /db_xref="taxon:3702"
    /map="unknown"
    /chromosome="IV"
    /name="unknown"
    /clone="F6N23"
    complement(187..907)
    /gene="F6N23.23"
    complement(join(187..347,396..512,598..907))
    /gene="F6N23.23"
    /codon_start=1
    /evidence=not experimental
    /protein_id="AAC13624.2"
    /db_xref="GI:6382044"
    /translation="MENLSDPNDNDHQSPRIDNQSAVETPVYTMISDFVYPR
    TCSETSGFSDQIDETNSFCSEAFPCNPVLTESKSKCLSGLEMQNECLVQEISE
    PELETWKEKFAKLLIGEDMSGSGKGVCTAVTISNAITNLVDCNKDTSFONVCSYVW
    TESEVRVNRRESIVERNELSFICMLHCRHP"
    complement(4919..7497)
    /gene="F6N23.22"
    complement(join(4919..4993,5078..5488,5567..6055,
    6157..6213,6498..6594,6935..7222,7349..7497))
    /gene="F6N23.22"
    /codon_start=1
    /evidence=not experimental
    /protein_id="AAC13623.1"
    /db_xref="GI:3047112"
    /translation="MSLWADGVEAAGRSQNSLLRKQLALAVRQVMSYAFWSS
    SUTQGVLEWGGYGNMRRKKSYESHYKYGLOKSKELRKLKYLMSLEDGSGTST
    THDNLNDNDNDNSKMLSPDLSDEWYLVMSYVFSQCLPGRASATGETTTL
    CNAQNAKLFKRSLLARTVCFYVIGVIELGTETISDNLRLIKKSLCMEISAH
    QNDDEKWEIKISEKHQLPLGISDELLHKRTISTVLNYSADRSGKNDKNTRRQRP
    NIVTSEPGKSLRMKQSQGVQVKKSONVLRKILHDVLMHTRKMPSPNSGLN
    QDDSRKKEKEFSLVMTPTNEVDKESILNNTIKYLQLEEARVEELESOMGSVN
    FVERQKTNLNDVLIETFSNGYDDSTKIDNSGETEQVTVPDKTHLRKVLKETE
    VVIEVRCSDRVIVADIMETLSNLDHDAFSPSRHTLNKFLNLNKKAFRGAASVSGM
    IKRELRRVIGDLF"
    10235..12675
    /gene="F6N23.1"
    join(10235..10495,10693..10896,10998..11192,11341..11505,

```

gene

CDS

gene

CDS

gene

CDS

gene

CDS

gene

```

11633..11893,11998..12216,12315..12534,12617..12675)
/gene="F6N23.1"
/Note="similar to the family of glycosyl hydrolases"
/evidence=not experimental
/codon_start=1
/db_xref="GI:3047123"
/translation="MAIRLNHSVIPYVKLGAPTRVSARSSLPFVSGDWGVSFTSGA
RPLVLAKVRAESTEDRVIDDDDDSTQLVSVFHLDDSLGLVIMNSEVVEPEE
LLDLRLTKSVNVDGVMDCWMIQVSHSTPOVNVWSGYKLFQMIRELGLKIQVWMSF
HECGNVGDDVHIOIPWREIGSQSPDIYFTSAGRNTTECLTWIDKQRLRGHTA
LEVTFMRSFRVEFFEFKEIPEIEVFGPGGELRIPSPYPAQGWKPIGIGERQOC
YKYLMSLKEAAEVRGHSFWGPGONTYNTSTPHGTGFRDGGDYDSYGRFLNWW
YSRVLIDHGRVLAMANLAFEGTCIAKLGSHWYKTAHAAELTAGNSNDRDGY
GPIAMFKKHDAALNFTCVELRTIDQHDPEALADPEGLVQVLAANDASIPVASE
NALPCYREGYNKILENAKPLTDPGRHLSCTFLRLNPTLMESQNFKEFERFLKMH
GEAVPDGLAPGTOETNPE"
complement(13027..14791)
/gene="F6N23.21"
complement(join(13027..13335,13480..13615,13712..13840,
13943..14001,14087..14791))
/gene="F6N23.21"
/codon_start=1
/evidence=not experimental
/protein_id="AAC13622.1"
/db_xref="GI:3047111"
/translation="MSILCCVFLVCYVCLGCTHNLWKKLYSAGHESNWLATSD
PEPTPRCRLIAYEENEDHPWAPDGYGIDPNHVLKKDYDQTEGRTVPMIYLD
HENGVDVLAIRGLNAKECDYAVLLKQTKDFDGYVHNGLLKAAMVFEEHVYL
RELLEANSVSLTFVSGSLAGVSVLLVFIQNRVRLGNIRKIRCAIFAIPRQMS
LHLAVTADVINSVLQDDLPRTTVALENVFKSIICLCLCLCTCKDTTFEBEKL
KDARLAPGELHYIVRKRLGRYPVTVTAVPVDSGEQIVLSGNATADHAITWI
BRESQRALDKIVROKSIDEDHEIRAAIMKASLNIWMSFSGSYGTDFHTEGESSA
SGMEGSPSGMGSRRKRWQDFIDCHFPVNDNSEHMIKFNQESQALLCQET"
15539..16687
/gene="F6N23.2"
join(15539..15554,15672..15764,15953..16026,16133..16208,
16307..16353,16466..16687)
/gene="F6N23.2"
/Note="Similar to protein kinase; coded for by A. thaliana
cDNA H36947; coded for by A. thaliana cDNA H37158"
/codon_start=1
/protein_id="AAC13612.1"
/db_xref="GI:3047101"
/translation="MNTESVVEFLGNVLLQRLPSSSLKEISEVVFKGVDRGDYVVR
ENQVDSGYELLOQVRAQVLRSGENYQEPFLKRYDPFGHGFQGVTSADVAYTE
LTCLLMSDRHALLKEIKSVSDSKERCLVEDILYLEPLDVRVSPNCPSPSSISVLMIK
VLTTHFYHHSLESEL"
16976..19151
/gene="F6N23.3"
join(16976..17111,17262..17415,17785..17889,18316..18417,
18618..18666,19056..19151)
/gene="F6N23.3"
/codon_start=1
/evidence=not experimental
/protein_id="AAC13635.1"
/db_xref="GI:3047124"
/translation="WEIYSNHCDDTVYASOALAAAKSVETMVKIHFHFCYFLLVGD
NIPILYDNRLDGNNFATRSVDAROKGTIFTLFASFOVSCYIPRIEHSVLDYRNO
VATISVFPFIDIRFCFEPNRSSTKQNSPPRCVYAFASDLIETATSLNPHREGMSVA
LSLDHSMWFRHPPVADDDLLFVIVSPATSESGFA'GKMFNRKGEVRCCIEE"
complement(20843..21366)
/gene="F6N23.20"
complement(join(20843..21002,21056,21103..21155,
21292..21366))
/gene="F6N23.20"
/codon_start=1
/evidence=not experimental
/protein_id="AAC13621.1"
/db_xref="GI:3047110"
/translation="MVATAVASVFLLTGTDYGPFPVNLDPYLSSSMILLVLEVLIRMF
SYWVWLHLFVIEHWIKKILAAQDSVQKFI'GPKGEPGKPSDTAK"
complement(22410..24753)

```

Db 1297 ATTAAGATGAGCGGAGACGGCTGTTGTTGGAGGAAAGTTGATGACTGAAGA 1356
 Qy 261 GlyLeuAlaAlaGluLysMetAspAlaArgGlyLeuLeuValAlaCysPhe 280
 Db 1357 GGATTAGCTGGCTGAGAAATGGATGCAAGGGTTGCTTTACTAGTCTCTGTTT 1416
 Qy 281 GlyValProSerPheArgSerThrAspLeuLeuAspLeuIleArgMetSerGlySer 300
 Db 1417 GGTGTTCTCTCAAACTTTAGAGATCAGATTCTGCTGATTTGATAGGATGAGTGGTTCG 1476
 Qy 301 AsnGluLeuAlaGlyAlaLeuLysArgSerGlnPheLeuValProMetValSer 318
 Db 1477 AATGAGATGCGCGTGCTTTGAAGCGGTACAGTTTCTTGCTCCCTATGCTCTCAGGTACC 1536
 Qy 318 ----- 318
 Db 1537 ATATCTGTTCTACTCGGTGAATTTCAITGCAAGGTGGTTCCTTTGTTGACATCATC 1596
 Qy 318 ----- 318
 Db 1597 GACCAACATCAAGTTCATCTTTGTTTTCGATAAGCTTGATGGTATAAACTAGGAGAGC 1656
 Qy 318 ----- 318
 Db 1657 ACATCAAAATTTAGAGTGCAATGACTGATGAGCCAAATCCCTAGTGAATAAATTAATCTG 1716
 Qy 318 ----- 318
 Db 1717 GAAAGAACTTGGAACTCTCAACCATAGTTTTCGTACGAAATTTGTTGCTTGTGCAAGACCA 1776
 Qy 318 ----- 318
 Db 1777 AATGATAGGCTATTGCTTGAAATAGTGTTCCTGTTGTTGTTTCCAAATTTGGAAGTAAATA 1836
 Qy 318 ----- 318
 Db 1837 TCATATGACTTAGCTGTGGTACTAATTAAGCTTAAGCAATGCCAACTCTAAGAAGTGG 1896
 Qy 319 ----- GlyLeuValGluSerSerIleLysArgGlyMet 329
 Db 1897 TACTTACACAATATTCCTATTGTTGGTCATGATGATAGTTGAATCAAGTATCAAGCGTGAATG 1956
 Qy 330 HisIleGluAlaLeuGluMetValThrPheGlyMetGluAspLysPheSerAlaAla 349
 Db 1957 CATATTGAGCTCTTGAGATGGTTTATACCTTTGGCATGGAGATAAGTTTTCAGTCTCT 2016
 Qy 350 LeuValLeuThrSerPheLeuLysMetSerLysGluSerPheGluArgAlaLysArgLys 369
 Db 2017 CTAGTTCTTAATCTTCTTAAAGATGAGCAAGGAGTCAATTTGAGAGGCGCAAAACGGAAA 2076
 Qy 370 AlaGlnSerProLeuAlaPheLys----- 377
 Db 2077 GCCCAGTCACCGCTGGCATTTGT-ATGAACCTTCCCTTGCACATTTATGACTCTTATGA 2135
 Qy 378 -----GluAla 379
 Db 2136 ACTCTTTATCATCTCATCTGAGCTGACCATTTGATATATTTATTTCTCAACAGCAAGAGCG 2195
 Qy 380 AlaThrLysGlnLeuAlaValLeuSerSerValMetGlnCysMetGluThrHisLysLeu 399
 Db 2196 GCTCAAGCAGCTAGCTGTGTTATCATCATGATGATGATGATGATGATGATGATGATGATGAT 2255
 Qy 400 AspProAlaLysGluLeuProGlyTrpGlnIleLysGluGlnIleValSerLeuGluLys 419
 Db 2256 GATCCTGGAAGAACTTACCAGGATGTCAGATCAAGAGCAAAATTTGATGTTGGAGAA 2315
 Qy 420 AspThrLeuGlnLeuAspLysGluMetGluGluLysAlaArgSerLeuLeuMetGlu 439
 Db 2316 GACACTCTTCAGCTCGCAAGAGATGCAAGAGAAAGCAAGATCTCTCAGTTTAATGGAG 2375
 Qy 440 GluAlaAlaLeuAlaLysArgMetTyrAsnGlnGlnIleLysArgProArgLeuSerPro 459
 Db 2376 GAAGCCGCACTTTCAGAGAGATGATATAACACAGATAAAAGTCCAGGTTGTACCC 2435

Qy 460 MetGluMetProProValThrSerSerTyrSerProIleTyrArgAspArgSerPhe 479
 Db 2436 ATGAAATGCCACAGTAACTTCTTCATCGTATTTCTCTATCTACCGGTAGTAGAGCTTT 2495
 Qy 480 ProSerGlnArgAspAspGlnAspGluLeuLeuValSerSerTyrLeuGly 499
 Db 2496 CCTAGTCAAAGAGACGATGACCAAGATGAAATATCAGCTCTTGTGAGTAGTACTCTCGGC 2555
 Qy 500 ProSerThrSerPheProHisArgSerArgSerProGluTyrMetValProLeuPro 519
 Db 2556 CCGTCAACATCTTTCTCATCGTCAAGAGATCCCCGGAATATATGTTCCACTTCCA 2615
 Qy 520 HisGlyGlyLeuGlyArgSerValTyrAlaTyrGluHisLeuAlaProAsnSerTyrSer 539
 Db 2616 CATGGTGGTTAGGAAGAAGTATATCATATGAACATCTGGCCCCAAATTCATCTCT 2675
 Qy 540 ProGlyHisGlyHisArgLeuHisArgGlnTyrSerProSerLeuValHisGlyGluArg 559
 Db 2676 CCAAGTCCACGACATAGACTTCTCCATTCATGGACACACAGATTACCATATGTTATACAA 2735
 Qy 560 HisProLeuGlnTyrSerProProIleHisGlyGlnGlnLeuProTyrTyrGlyLeuGln 579
 Db 2736 CATCCACTACGACTCTCTCCATTCATGGACACACAGATTACCATATGTTATACAA 2795
 Qy 580 ArgValTyrArgHisSerProSerGluGluArgTyrLeuGlyLeuSerAsnGlnArgSer 599
 Db 2796 AGGGTTTACAGACATTCACCATCTGAAGAAAGATATTTGGGTTTATCCAAATCAAAGGTCT 2855
 Qy 600 ProArgSerAsnSerSerLeuAspProLys 609
 Db 2856 CTTCCGAGTAACTCATCATTAGACCCCA 2885
 RESULT 10
 F6N23
 LOCUS 91849 bp DNA linear PLN 12-NOV-1999
 DEFINITION Arabidopsis thaliana BAC F6N23.
 ACCESSION AF058919
 VERSION AF058919.2 GI:6382042
 KEYWORDS
 SOURCE Arabidopsis thaliana (thale cress)
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
 REFERENCE 1 (bases 1 to 91849)
 Geisel, C.
 The sequence of A. thaliana F6N23
 Unpublished
 REFERENCE 2 (bases 1 to 91849)
 The A. thaliana Genome Sequencing Project
 Unpublished
 REFERENCE 3 (bases 1 to 91849)
 Waterston, R.
 Direct Submission
 Submitted (10-APR-1998) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
 REFERENCE 4 (bases 1 to 91849)
 Waterston, R.
 Direct Submission
 Submitted (12-NOV-1999) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
 On Nov 12, 1999 this sequence version replaced gi:3047100.
 COMMENT Submitted by:
 Genome Sequencing Center
 Department of Genetics, Washington University,
 St. Louis, MO 63108, USA
 e-mail: rwaterston@wustl.edu

MAPPING: Clones were assigned to the YAC map by hybridization by M. Lodhi, Cold Spring Harbor Laboratories, and fingerprinted

Db 2086 ACTCTTTATCATCATCTGAGTCTGACCATTCATATATTTATTTCTCAACAGAAAGAGCG 2145

QY 380 AlaThrLysGluLeuAlaValLysSerValMetGlnCysMetGluThrHisLysLeu 399

Db 2146 GCTACAAGACGCTAGCTGTGTATCATCAGTATGTCAGTGTATGGAGACTCAAGTTA 2205

QY 400 AspProAlaLysGluLeuProGlyTrpGlnIleLysGluGlnIleValSerLeuGluLys 419

Db 2206 GATCTCTCGAAAGAACTACCAAGATGGCAGATCAAGAGCAAAATTTAGTCTGGAGAAA 2265

QY 420 AspThrLeuGlnLeuAspLysGluMetGluGluLysAlaArgSerLeuMetGlu 439

Db 2266 GACACTCTTCAGCTCGAACAAGATGAAGAGAAAGCAAGATCTTCAGTTTAATGGAG 2325

QY 440 GluAlaLeuAlaLysArgMetTyrAsnGlnGlnIleLysArgProArgLeuSerPro 459

Db 2326 GAAGCGCACTTGGCAAGAAATGATATCAACACAGATAAAACGTCCAAGTTGTCAACC 2385

QY 460 MetGluMetProProValThrSerSerTyrSerProIleTyrArgAspArgSerPhe 479

Db 2386 ATGGAAATGCCACCAAGTAACTTCITCATCGTATTTCTCTATCTACCGTGATAGAAGCTTT 2445

QY 480 ProSerGlnArgAspAspGlnAspGluIleSerAlaLeuValSerTyrLeuGly 499

Db 2446 CTAGTCAAGAGACGATGACCAAGATGAATATACAGCTCTTGAGTAGTACTTCGGC 2505

QY 500 ProSerThrSerPheHisArgSerArgArgSerProGluTyrMetValProLeuPro 519

Db 2506 CGGTCAACATCTTTTCTCATCGCTCAAGAAAGATCCCGGAATATATGGTTCCACTTCCA 2565

QY 520 HisGlyGlyLeuGlyArgSerValTyrAlaTyrGluHisLeuAlaProAsnSerTyrSer 539

Db 2566 CATGGTGGTTAGGAGAGTATATCATATGACATCTGGCCCCAATTCATCTCT 2625

QY 540 ProGlyHisGlyHisArgLeuHisArgGlnTyrSerProSerLeuValHisGlyGlnArg 559

Db 2626 CCAGGTCAACGACATAGACTTCATCGACAGTACTCTCGTCTTTGGTTCCGGACAGAGA 2685

QY 560 HisProLeuGlnTyrSerProProIleHisGlyGlnGlnGlnLeuProTyrGlyIleGln 579

Db 2686 CATCCACATACAGTACTCTCTCCCAATTCATGACAAACAGATTTACCATATGGTATACAA 2745

QY 580 ArgValTyrArgHisSerProSerGluGluArgTyrLeuGlyLeuSerAsnGlnArgSer 599

Db 2746 AGGGTTTACAGACATTCACCATCTGACAAAGAGATTTTGGTTTATCCATCAAGGTCT 2805

QY 600 ProArgSerAsnSerSerLeuAspProLys 609

Db 2806 CCTCGCAGTAACTCATATTAGACCCCAAA 2835

RESULT 9

AY198404 3758 bp DNA linear PLN 12-JUN-2003

LOCUS Arabidopsis thaliana ecotype Cvi FRIGIDA-like protein gene,

DEFINITION complete sequence.

ACCESSION AY198404

VERSION AY198404.1 GI:31558916

KEYWORDS Arabidopsis thaliana (chale cress)

SOURCE Arabidopsis thaliana

ORGANISM Arabidopsis thaliana

REFERENCE 1 (bases 1 to 3758)

AUTHORS Gazzani, S., Gendall, A.R., Lister, C. and Dean, C.

TITLE Analysis of the molecular basis of flowering time variation in Arabidopsis accessions

JOURNAL Plant Physiol. 132 (2), 1107-1114 (2003)

REFERENCE 2 (bases 1 to 3758)

AUTHORS Gazzani, S. and Dean, C.

TITLE Direct Submission

JOURNAL Submitted (17-DEC-2002) Cell and Developmental Biology, John Innes Center, Colney Lane, Norwich NR4 7UH, UK

FEATURES

Source 1..3758

Location/Qualifiers

/organism="Arabidopsis thaliana"

/mol_type="genomic DNA"

/db_xref="taxon:3702"

/chromosome="4"

/map="between mi51 and mi122"

/note="ecotype: Cvi"

misc_feature 577..2888

/note="similar to FRIGIDA (FRI)"

BASE COUNT 1178 a 714 c 774 g 1092 t

ORIGIN

Alignment Scores:

Pred. No.: 7,09e-185 Length: 3758

Score: 2950,00 Matches: 604

Percent Similarity: 78,57% Conservative: 1

Best Local Similarity: 78,44% Mismatches: 4

Query Match: 95,04% Indels: 162

DB: 8 Gaps: 2

US-09-890-475-1 (1-609) x AY198404 (1-3758)

QY 1 MetSerAsnTyrProProThrValAlaAlaGlnProThrThrThrAlaAsnProLeuLeu 20

Db 577 ATGTCCAATTAATCACCAGCGGTGGCGCGCAACCAACAGCGCGAATCCACTGCTG 636

QY 21 GlnArgHisGlnSerGluGlnArgArgGluLeuProLysIleValGluThrGluSer 40

Db 637 CAGCGACATCAATCTGAACAGCGAGAGAAATTAACGAGATTCTCGAACAAGATCT 696

QY 41 ThrSerMetAspIleThrIleGlyGlnSerLysGlnProGlnPheLeuLysSerIleAsp 60

Db 697 ACAAGTAGGACATTAACGATCGGTCAATCTTAAGCAGCTCAATTTTGAATCCATAGAC 756

QY 61 GluLeuAlaAlaPheSerValAlaValGluThrPheLysArgGlnPheAspLeuGln 80

Db 757 GAATTAGCTGGTTCAGTTGAGTGGAAACATTCAATGCCAATTCGATGATCTTCAG 816

QY 81 LysHisIleGluSerIleGluAsnAlaIleAspSerLysLeuGluSerAsnGlyValVal 100

Db 817 AAGCACATCGAGTCAATCGAAACGCAATTAATTCGAGAGTAACCGCGTTGTC 876

QY 101 LeuAlaAlaArgAsnAsnAsnPheHisGlnProMetLeuSerProProArgAsnVal 120

Db 877 CTCGCGCGCGGAGCAATATTTCCATCAGCCGATGTATCGCTCCGCGGAACAATGA 936

QY 121 SerValGluThrThrValThrValSerGlnProSerGlnGluIleValProGluThrSer 140

Db 937 TCTGTAGAAACCCAGCTCACTGTGAGCCAAACCGCTCTCAGGAGATTGTACCGAGACGTG 996

QY 141 AsnLysProGluGlyGlyArgMetCysGluLeuMetCysSerLysGlyLeuArgLysTyr 160

Db 997 AATAAACCGGAGGGGAGCGATGTGTGAGTTGATGTAGCAAGGTCTGCGTAATATAC 1056

QY 161 IleTyrAlaAsnIleSerAspGlnAlaLysLeuMetGluGluIleProSerAlaLeuLys 180

Db 1057 ATATAGCGAATATCTCTGAACAGCTAAGTTAATGGAAGATTCTCTCAGCTTTGAAA 1116

QY 181 LeuAlaLysGluProAlaLysPheValLeuAspCysIleGlyLysPheTyrLeuGlnGly 200

Db 1117 TTGCCAAGGAGCGAGCAAGTTGTATGAGTTGATTTGGCAAGTTTACTTACAAAGG 1176

QY 201 ArgArgAlaPheThrLysGluSerProMetSerSerAlaArgGlnValSerLeuLeuIle 220

Db 1177 CGTAGACATTTACTAAGAGTGGCTATGAGCTCTGCGAGACAAGTTTCGCTTCTATA 1236

QY 221 LeuGluSerPheLeuLeuMetProAspArgGlyLysGlyLysValLysIleGluSerTrp 240

Db 1237 CTGGAGTCTTTCTTCTAATCCCTGATCGTGGTTAAGGGAAGGTGAAGATTGAGAGTTGG 1296

QY 241 IleLysAspGluAlaGluThrAlaAlaValAlaTTPArgLysArgLeuMetThrGluGly 260

1007	Db	ATATACGGCAATATCTCTGAACAAGCTAAAGTTAAATGGAAGAGATTCCCTTACGCTTGGAAA	1066
181	QY	LeuAlaLysGluProAlaLysPheValLeuAspCysIleGlyLysPheTyrLeuGlnGly	200
1067	Db	TGGGCCAAGAGACGACGAAGTTTGATTGGATTGTATTGGCAGATTTTACTTTCAAGGG	1126
201	QY	ArgArgAlaPheThrLysGluSerProMetSerSerAlaArgGlnValSerLeuLeuIle	220
1127	Db	CGTAGAGCATTTACTAAAGAGTCGCTATGAGCTCTGCAGACAAGTTTCGCTCTTTATA	1186
221	QY	LeuGluSerPheLeuLeuMetProAspArgGlyLysGlyLysValLysIleGluSerTyr	240
1187	Db	CTGGAGTCTTTCTCTAAATGCTGATCTGTGTGAATGCTGTGAAGGAAGTGAAGATTGAGAGTTGG	1246
241	QY	IleLysAspGluAlaGluThrAlaAlaValAlaTyrPArgLysArgLeuMetThrGluGly	260
1247	Db	ATTAAAGATGAGCGGAGACGGCTGCTGTGTCTTGAGGAAAAGGTTGATGACTGAAGGA	1306
261	QY	GlyLeuAlaAlaGluLysMetAspAlaArgGlyLeuLeuLeuLeuValAlaCysPhe	280
1307	Db	GGATTAGCTCGGCTGAGAAAATGGATGCAAGGGTTGCTTTTACTAGTTGCTTGTGTTTT	1366
281	QY	GlyValProSerAsnPheArgSerThrAspLeuLeuAspLeuIleArgMetSerGlySer	300
1367	Db	GGTGTTCTCTCAAACTTTAGAGTACAGATTTCGTGGATTGATAAGATGAGTGGTTCG	1426
301	QY	AsnGluIleAlaGlyAlaLeuLysArgSerGlnPheLeuValProMetValSer----	318
1427	Db	AATGAGATTGCGGTGCTTTGAAAGCGGTACACAGTTTCTGTCCCTATGGTCTCAGGTACC	1486
318	QY	-----	318
1487	Db	ATATTCTGTTCTCACTCGGTGAATTTTCATTGCAAAAGGTGGTTTCCTTTGTTGACATCATC	1546
318	QY	-----	318
1547	Db	GACCAACATCAAGTTCCATCTTTTGTGTTTCGATAAGCTTGATGGTATAAACTAGGAGAGC	1606
318	QY	-----	318
1607	Db	ACATCAAAATTTAGAGTGCATGACTGATGAGCCAAATCCTAGCTAGAGAAATTAATCTG	1666
318	QY	-----	318
1667	Db	GAAGAACTTGGAACTCTCAACCATAGGTTTGGTACGAAATTTGTTGTTGCAGAACCA	1726
318	QY	-----	318
1727	Db	AATGATAGGCTATTGCGCTGAAATAGTGTTCCTGTGTTTCCAAATATTTGGAAGTTAAAA	1786
318	QY	-----	318
1787	Db	TCATATGACTTAGTGTTGGATACTAATTAAGCTTTAAGCAATGCCAACTCTAAGAGTGG	1846
319	QY	-----GlyIleValGluSerSerIleLysArgGlyMet	329
1847	Db	TACTTACACAATATTCATTGGTTCATAGTATAGTTGAATCAAGATATCAAGCGTGGAAATG	1906
330	QY	HisIleGluAlaLeuGluMetValTyrThrPheGlyMetGluAspLysPheSerAlaAla	349
1907	Db	CATATTGAAGCTCTTGAGATGGTTTATACCTTTGCGATGGAGGATAGTTTTTTCAGTGCT	1966
350	QY	LeuValLeuThrSerPheLeuLysMetSerLysGluSerPheGluArgAlaLysArgLys	369
1967	Db	CTAGTTCTTCACTTCATTCTTAAAGATGACGAAGGAGTCATTTGAGAGGGCAAAACGAAA	2026
370	QY	AlaGlnSerProLeuAlaPheLys-----	377
2027	Db	GCCAGTCACCGTGGCATTTGT-ATGAACCCCTCCCTTGCACATTAATGATACCTTTATGA	2085
378	QY	-----GluAla	379

```

Db      1114 TTGGCCAGAGCCAGCGAGTTGTTATTTGGATTGTATGGCAAGTTTTTACTTACAGGG 1173
QY      201 ArgArgAlaPheThrLysGluSerProMetSerSerAlaArgGlnValSerLeuLeuIle 220
Db      1174 CGTAGAGCATTTACTAAAGAGTGGCTATGAGCTCGCAGACAAGTTTCGCTTCATTATA 1233
QY      221 LeuGluSerPheLeuLeuMetProAspArgGlyLysGlyLysValIleGluSerTrp 240
Db      1234 CTGGAGTCTTTTCTTCTAATGCTGATCGTGGTAAGGAAGGTGAAGATTGAGAGTTGG 1293
QY      241 IleLysAspGluAlaGluThrAlaAlaValAlaIleArgLysArgLeuMetThrGluGly 260
Db      1294 ATTAAGATGAGCGCGAGACGCGTCTGCTGCTTGGAGGAAGAGTTGATGACTGAAGA 1353
QY      261 GlyLeuAlaAlaAlaGluLysMetAspAlaArgGlyLeuLeuLeuValAlaCysPhe 280
Db      1354 GGATTAGCTGGCGCTGGAATAATGGATGCAAGGGGTTTGCTTTTACTAGTTGCTTTGTTT 1413
QY      281 GlyValProSerAsnPheArgSerThrAspLeuLeuAspLeuIleArgMetSerGlySer 300
Db      1414 GGTGTTCTTCAAACCTTTAGGAGTACAGATTGCTGGATTGTGATAAGGATGAGTGTGCTG 1473
QY      301 AsnGluIleAlaGlyAlaLeuLysArgSerGlnPheLeuValProMetValSer----- 318
Db      1474 AATGAGATTGGCGGTGCTTTGAAGCGGTACAGTTTCTGTGCTATGCTCAGGTACC 1533
QY      318 ----- 318
Db      1534 ATATTCTGTTCTACCTCGGTGANTTCATTGCAAGGTGGTTCCTTTTGTGTGACATCATC 1593
QY      318 ----- 318
Db      1594 GACCAACATCAAGTTCCATCTTTGTTTTTCGATAAGCTTGATGTTATATAAAGTAGGAGAGC 1653
QY      318 ----- 318
Db      1654 ACATCAAAATTTAGAGTGCATGACTGATTGAGCCAAATCCTAGCTAGAAAATTAATCTG 1713
QY      318 ----- 318
Db      1714 GAAAGAACTTGGAACTCTCAACCATAGGTTTTTGGTACGAAATTTGCTGTGTCAGAACCA 1773
QY      318 ----- 318
Db      1774 AATGATAGCTATTGCTTGGCTTGAATAGTGTTCCTTGTGTTTCCAAATATCGAAGTAAAA 1833
QY      318 ----- 318
Db      1834 TCATATGACTTAGCTGTGGATACATACTAATTAAGCTTAAGCAATGCCAATCTAAGAAGTGG 1893
QY      319 -----GlyIleValGluSerSerIleLysArgGlyMet 329
Db      1894 TACTTACACAAATTTCTATTGTTGTCATAGGTATAGTTGAATCAAGATATCAAGCGTGAATG 1953
QY      330 HisIleGluAlaLeuGluMetValThrPheGlyMetGluAspLysPheSerAlaAla 349
Db      1954 CATATTGAGCTCTTGAGATGGTTTATACCTTTGCGATGGAGGATAAGTTTTCAGTGTCT 2013
QY      350 LeuValLeuThrSerPheLeuLysMetSerLysGluSerPheGluArgAlaLysArgLys 369
Db      2014 CTAGTTCTTAACCTTCTTAAGATGAGCAAGGAGTCAITTGAGAGGGCAAAACGGAAA 2073
QY      370 AlaGlnSerProLeuAlaPheLys----- 377
Db      2074 GCCAGTACCCTGGCAATTTGT-ATGAAACCTTCCTTGCACATATATGACCTTTATGA 2132
QY      378 -----GluAla 379
Db      2133 ACTCTTTATCATCTGAGTCTGACCATTCATATATTTTCTCAACAGAAAGAGCG 2192
QY      380 AlaThrLysGlnLeuAlaValLeuSerSerValMetGlnCysMetGluThrHisLysLeu 399

```

```

Db      2193 GCTCAAAAGAGCAGCTAGCTGTGTTTATCATCATGATTATGCAGTGTATGAGACTCACAAGTTA 2252
QY      400 AspProLalysGluLeuProGlyTTPGlnIleLysGluGlnIleValSerLeuGluLys 419
Db      2253 GATCTCGAAAGAACTTACCAGGATGGCAGATCAAAAGAGCAAAATTTAGCTTGGAGAAA 2312
QY      420 AspThrLeuGlnLeuAspLysGluMetGluGluLysAlaArgSerLeuSerLeuMetGlu 439
Db      2313 GACACTCTTCAGCTCGAAGAGAGATGGAAGAGAAAGCAAGATCTCTCAGTTTAAATGGAG 2372
QY      440 GluAlaAlaLeuAlaLysArgMetTyrAsnGlnGlnIleLysArgProArgLeuSerPro 459
Db      2373 GAAGCGCACCTTGCCAAAGAGATGTATTAACCAACAGATAAAACGTCCAAGTTGTCAACC 2432
QY      460 MetGluMetProProValThrSerSerTyrSerProIleTyrArgAspArgSerPhe 479
Db      2433 ATGGAATGTCACCAAGTAACCTTCTCATCGTATTTCTCTATCTTACCGTGATAGAAGCTTT 2492
QY      480 ProSerGlnArgAspAspGlnAspGluIleSerAlaLeuValSerSerTyrLeuGly 499
Db      2493 CCTAGTCAAAGAGACGATGACCAAGATGAATATCAGCTCTTGTGAGTAGTTACCTCGGC 2552
QY      500 ProSerThrSerPheProHisArgSerArgArgSerProGluTyrMetValProLeuPro 519
Db      2553 CCGTCAACATCTTTTCTCATCGCTCAAGAAGATCCCGGAATATATGTTCCACTTCCA 2612
QY      520 HisGlyGlyLeuGlyArgSerValTyrAlaTyrGluHisLeuAlaProAsnSerTyrSer 539
Db      2613 CATGGTGGTTAGGAAGAGTGTATATGATATGAACATCTGCCCCCAATTCATACTCT 2672
QY      540 ProGlyHisGlyHisArgLeuHisArgGlnTyrSerProSerLeuValHisGlyGlnArg 559
Db      2673 CCAGGTCACGACATAGACTTCATCGACAGTACTCTCCGTCTTTGTTGTCACGACAGAGA 2732
QY      560 HisProLeuGlnTyrSerProIleHisGlyGlnGlnLeuProTyrGlyIleGln 579
Db      2733 CATCACTACAGTACTCTCTCCAAATTCATGGAACAACAGTACCATTACATATGTTATCAA 2792
QY      580 ArgValTyrArgHisSerProSerGluGluArgTyrLeuGlyLeuSerAsnGlnArgSer 599
Db      2793 AGGGTTTACAGACATTCACCATCTGAAGAAGATATTGGGTTTATCCAAATCAAAGGTCT 2852
QY      600 ProArgSerAsnSerSerLeuAspProLys 609
Db      2853 CCTCGAGTAACATCATCTAGACCCCAAA 2882

RESULT 8
AV198405
LOCUS   Arabidopsis thaliana ecotype Wai-2 FRIGIDA (FRI) gene, complete
DEFINITION
ACCESSION   AV198405
VERSION     AV198405.1  GI:31558917
KEYWORDS
SOURCE      Arabidopsis thaliana (thale cress)
ORGANISM    Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
REFERENCE   1 (bases 1 to 3301)
AUTHORS     Gazzani, S., Gendall, A.R., Lister, C. and Dean, C.
TITLE       Analysis of the molecular basis of flowering time variation in
Arabidopsis accessions
JOURNAL     Plant Physiol. 132 (2), 1107-1114 (2003)
REFERENCE   2 (bases 1 to 3301)
AUTHORS     Gazzani, S. and Dean, C.
TITLE       Direct Submission
JOURNAL
SUBMITTED   (17-DEC-2002) Cell and Developmental Biology, John Innes
Center, Colney Lane, Norwich NR4 7UH, UK
FEATURES
     1..3301
         Location/Qualifiers
             /organism="Arabidopsis thaliana"
             /mol_type="genomic DNA"

```

```

Db      2184  GATCCTCGAAGAACTACACGAGGATGCCAGATCAAGACGAAATTTGATGCTTGAGAGAAA 2243
Qy      420  AspThrLeuGlnLeuAspLysGluMetGluGluLysAlaArgSerLeuSerLeuMetGlu 439
Db      2244  GACACTCTTCAGCTCGCAAGAGATGGAAGAGAAAGCAAGATCTCTCAGTTTATGAG 2303
Qy      440  GluAlaAlaLeuAlaLysArgMetTyrAsnGlnGlnLysArgProArgLeuSerPro 459
Db      2304  GAAGCCGACCTTGCCAAAGAAATGTATACCAACAGATAAACGTCACCAAGGTGTCACCC 2363
Qy      460  MetGluMetProValThrSerSerSerTyrSerProLeyrArgAspArgSerPhe 479
Db      2364  ATGGAATGCCACAGTAATCTTCTTCATCTATCTCTCCATCTACCGTGATAGAGCTTT 2423
Qy      480  ProSerGlnArgAspAspGlnAspGluLysSerAlaLeuValSerSerTyrLeuGly 499
Db      2424  CCTAGTCAAGAGACGATGACCAAGATGAATATACGCTCTTGAGTAGTTTACCTCGC 2483
Qy      500  ProSerThrSerPheProHisArgSerArgArgSerProGluTyrMetValProLeuPro 519
Db      2484  CCGTCAACATCTTTCTCATCGCTCAAGAGATCCCGGAATATATGTTCCACTTCCA 2543
Qy      520  HisGlyGlyLeuGlyArgSerValTyrAlaTyrGluHisLeuAlaProAsnSerTyrSer 539
Db      2544  CATGTCGGTGTAGGAAGATGTATATGCATATGAACATCTGCGCCCAATTCATCTCT 2603
Qy      540  ProGlyHisGlyHisArgLeuHisArgGlnTyrSerProSerLeuValHisGlyGlnArg 559
Db      2604  CCAGTTCAGGACATAGACTTCATCGACGATCTCTCCGCTCTTGGTTCACGACAGAGA 2663
Qy      560  HisProLeuGlnTyrSerProLeyrHisGlyGlnGlnGlnLeuProTyrGlyLeuGln 579
Db      2664  CATCACTACAGTACTCTCTCCATTCATGACAAACAGTTACCATATGATATACAA 2723
Qy      580  ArgValTyrArgHisSerProSerGluGluArgTyrLeuGluSerAsnGlnArgSer 599
Db      2724  AGGTTTACAGACATTCACCATCTGAAGAAGATATTTGGTTTATCCAAATTCAGGCT 2783
Qy      600  ProArgSerAsnSerSerLeuAspProLys 609
Db      2784  CCTCGCAGTAACATCATCTAGACCCCAAA 2813

RESULT 7
AY198401
LOCUS      AY198401          3739 bp    DNA    linear    PLN 12-JUN-2003
DEFINITION Arabidopsis thaliana ecotype Shakhdara FRIGIDA (FRI) gene, complete cds.
ACCESSION  AY198401
VERSION     AY198401.1  GI:31558910
KEYWORDS   Arabidopsis thaliana (thale cress)
SOURCE      Arabidopsis thaliana
ORGANISM    Arabidopsis thaliana
REFERENCE   1  (bases 1 to 3739)
AUTHORS     Gazzani, S., Gendall, A.R., Lister, C. and Dean, C.
TITLE       Analysis of the molecular basis of flowering time variation in Arabidopsis accessions
JOURNAL     Plant Physiol. 132 (2), 1107-1114 (2003)
REFERENCE   2  (bases 1 to 3739)
AUTHORS     Lister, C. and Dean, C.
TITLE       Direct Submission
JOURNAL     Submitted (17-DEC-2002) Cell and Developmental Biology, John Innes Center, Colney Lane, Norwich NR4 7UH, UK
FEATURES   Location/Qualifiers
            1..3739
                /organism="Arabidopsis thaliana"
                /mol_type="genomic DNA"
                /db_xref="taxon:3702"
                /chromosome="4"
                /map="between ml51 and ml122"
                /note="ecotype: Shakhdara"

```

```

gene
mRNA
CDS
BASE COUNT  1172 a 711 c 767 g 1089 t
ORIGIN
Alignment Scores:
Pred. No.:      1,59e-186      Length:      3739
Score:          2975.00      Matches:    607
Percent Similarity: 78.83%      Conservative: 0
Best Local Similarity: 78.83%      Mismatches: 2
Query Match:      95.84%      Indels:    162
DB:              8      Gaps:      2

US-09-890-475-1 (1-609) x AY198401 (1-3739)
Qy      1  MetSerAsnTyrProProThrValAlaAlaGlnProThrThrThrAlaAsnProLeuLeu 20
Db      574  ATGTCCAATTTATCCACCGACGGTGGCGGCAACCCACACGACGCGGAATCCCACTGCTG 633
Qy      21  GlnArgHisGlnSerGluGlnArgArgGluLeuProLysIleValGluThrGluSer 40
Db      634  CAGCGACATCAATCTGAACAGCGACGAGAGAGATTTACCGAGATTTGTCGAACAGAGTCT 693
Qy      41  ThrSerMetAspIleThrIleGlyGlnSerLysGlnProGlnPheLeuLysSerIleAsp 60
Db      694  ACAAGTATGACATTCAGATCGGTCAATCTAAGCAGCCTCAAAATTTGAAATTCATAGAC 753
Qy      61  GluLeuAlaAlaPheSerValAlaValGluThrPheLysArgGlnPheAspAspLeuGln 80
Db      754  GAATTAGTCGCTTTTCAGTTGAGTGAACATTTCAACCGCAATTCGATGATCTTCAG 813
Qy      81  LysHisIleGluSerIleGluAsnAlaIleAspSerLysLeuGluSerAsnGlyValVal 100
Db      814  AAGCACATCGAGTCAATCGAAACCAATTTGATTTCAAACTCGAGAGTAACGGCGTTGTC 873
Qy      101  LeuAlaAlaArgAsnAsnAsnPheHisGlnProMetLeuSerProProArgAsnAsnVal 120
Db      874  CTGCGCGCGGAGCAATAAATTTCCATCAGCGGATGTTATCGCTCCGCGGAACAATGTA 933
Qy      121  SerValGluThrThrValThrValSerGlnProSerGlnGluIleValProGluThrSer 140
Db      934  TCTGTAGAAACACCGCTCACTGTGAGCCAAACCGTCTCAGGAGATTTGTACCGGAGCGTCG 993
Qy      141  AsnLysProGluGlyGlyArgMetCysGluLeuMetCysSerLysGlyLeuArgLysTyr 160
Db      994  AATAAACCGGAGGGGCGATGTTGTGAGTTGATGTAGTAAAGGCTCGCGTAATAATAC 1053
Qy      161  IleTyrAlaAsnIleSerAspGlnAlaLysLeuMetGluGluIleProSerAlaLeuLys 180
Db      1054  ATATAGCGGAATATCTCTGATCAAGCTAAGTAAAGTAGAGAGATTCCTTCAGCTTTGAA 1113
Qy      181  LeuAlaLysGluProAlaLysPheValLeuAspCysIleGlyLysPheTyrLeuGlnGly 200

```

```
CDS
/product="FRIGIDA"
Join(505, .1459,1853, .2025,2115, .2816)
/ gene="PRI"
/ note="flowering gene"
/codon_start=1
/product="FRIGIDA"
/protein_id="AAP49808.1"
/db_xref="GI:31558913"
translation="MSNYPTVAQPTTANPLLRHQSRQRRELKPIVETSTSD
TRIGSKPOLKSIDELAAASVAVETFRQFDLQKHIESIENAIKSLKESNGVIA
ARNNPHQPMUSPPRNYSVETTVTSQSPQEIIVPTSNKPEGGRMCELMCSKGLRKY
IYANTSDAKLMEETPSALKAKPAKFLDCIKFYLOQRRAFTKESPMSSARQVSL
LLESLFLMPDRKGKVKIISWKIDAEATAVAVWRKRLMTGGGLAAEKMDARGLLIL
VACVFGNFRSTLDLLIRMSGNEIAJALKRSQFLVPMVSGIVESSIKRGWHIBAL
EMVYTFMEDKFSALVLTSLKMSKSPERAKKAQSLAPKFAATKOLAVLSSVMO
CMETHKLDPKELPWQJKEQIVLSLEKDTLQDKEMEEKARSLSIMEAALAKRMVQ
QIKRPLSFPMPVTSYSPTFRDQDFPSQDDDDQDEISALVSSLYKSPSTSPFPHRS
RRSPYVPLPHGGGYSVAYEHLAPNSYSPGHRLHRQYSPSLVHGHRHFLQYSP
PIHGQOOLPYGIQVRYRHSPEERYLGLSNQSRSPRSNSLDPK"
BASE COUNT 990 a 638 c 666 g 967 t
ORIGIN

Alignment Scores:
Pred. No.: 1,33e-186 Length: 3261
Score: 2975.00 Matches: 607
Percent Similarity: 78.83% Conservative: 0
Best Local Similarity: 78.83% Mismatches: 2
Query Match: 95.84% Indels: 162
DB: 8 Gaps: 2

US-09-890-475-1 (1-609) x AY198402 (1-3261)
QY 1 MetSerAsnTyrProProThrValAlaAlaGlnProThrThrAlaAsnProLeuLeu 20
Db 505 ATGTCATATTATCCACGAGCGTGGCGCGCAACCCACACGACGCGCAATCCACTGCTG 564
QY 21 GlnArgHisGlnSerGluGlnArgArgGlnLeuProLysIleValGluThrGluSer 40
Db 565 CAGCGACATCAATCTGAACGACGACGAGAGAAATTAACGAGATTTGCGAAACAGAGTCT 624
QY 41 ThrSerMetAspIleThrIleGlyGlnSerLysGlnProGlnPheLeuLysSerIleAsp 60
Db 625 ACAAGTATGGACATTACGATCGGTCAATCTAAGCAGCGCTCAATTTTGAATCCATAGAC 684
QY 61 GluLeuAlaAlaPheSerValAlaValGluThrPheLysArgGlnPheAspAspLeuGln 80
Db 685 GAATTAGTCTGCTTTTCAGTTGCGAGTGAACCAATTCGATTCGATGATGATGATGATG 744
QY 81 LysHisIleGluSerIleGluAsnAlaIleAspSerLysLeuGluSerAsnGlyValVal 100
Db 745 AAGCAGATCGAGTCAATCGAAAAACCAATTCATTCGAACTCGAGAGTAACGGCGTTGTC 804
QY 101 LeuAlaAlaArgAsnAsnAsnPheHisGlnProMetLeuSerProProArgAsnAsnVal 120
Db 805 CTCGCGCGCGCGGAAACAATAATTTCCATCAGCGATGTTATCGCGCTCGCGGAAATGTA 864
QY 121 SerValGluThrThrValThrValSerGlnProSerGlnGluIleValProGluThrSer 140
Db 865 TCTGTAGAAACACCGCTCAGTGTAGCCCAACCGTCTCAGGAGATTGTACCGGAGACGTG 924
QY 141 AsnLysProGluGlyGlyArgMetCysGluLeuMetCysSerLysGlyLeuArgLysTyr 160
Db 925 AATAAACCGAGGGGGGAGATGTGTGTGAGTTGATGTGTAGCAAGGTCGCGTAAATAC 984
QY 161 IleTyrAlaAsnIleSerAspGlnAlaLysLeuMetGluGluIleProSerAlaLeuLys 180
Db 985 ATATACCGGAATATCTCGATCAAGCTAAGTTAATGAAGAGATTCCTTCAGGTTTGA 1044
QY 181 LeuAlaLysGluProAlaLysPheValLeuAspCysIleGlyLysPheTyrLeuGlnGly 200
Db 1045 TTGGCCAAAGGACCGACGAGATTGTTGTTGATGTTGATGTTGATGTTGATGTTGATG 1104
QY 201 ArgArgAlaPheThrLysGluSerProMetSerSerAlaArgGlnValSerLeuLeuIle 220
```

```
Db 1105 CGTAGAGCATTTACTAAAGAGTCGCTATGAGCTCTCGAGACAAAGTTTCGCTCTTTATA 1164
QY 221 LeuGluSerPheLeuLeuMetProAspArgGlyLysGlyLysValLysIleGluSerTyr 240
Db 1165 CTGGAGTCTTTCTTCTTAATGCCTGATCGTGGTAAAGGAGGTGAAGATTGAGAGTTGG 1224
QY 241 IleLysAspGluAlaGluThrAlaAlaValAlaIleThrPheGlyArgLeuMetThrGluGly 260
Db 1255 ATTAAGAGATAGGCGGAGACGCTGCTGTGCTGGAGGAAAAGTTGATGACTGAAGGA 1284
QY 261 GlyLeuAlaAlaAlaGluLysMetAspAlaArgGlyLeuLeuLeuValAlaCysPhe 280
Db 1285 GGAATTAGCTCGGCTGAGAAAATGATGCAAGGGTTGCTTTACTAGTTGCTGTTT 1344
QY 281 GlyValProSerAsnPheArgSerThrAspLeuLeuAspLeuIleArgMetSerGlySer 300
Db 1345 GGTGTCTCTCAAACTTTAGGAGTACAGATTGCTGTGATTGTAAGGATGAGTGGTTTCG 1404
QY 301 AsnGluIleAlaGlyAlaLeuLysArgSerGlnPheLeuValProMetValSer 318
Db 1405 AATGAGATTGCGGTGCTTTGAAGCGGTACAGTTTCTTCTCCCTATGGTCTCAGGTACC 1464
QY 318 ----- 318
Db 1465 ATATTCTGTTCTCAGTCGGTGAATTTCAATTGCAGAGGTGCTTCTTTTGTGACATCATC 1524
QY 318 ----- 318
Db 1525 GACCAACATCAAGTTCCATCTTTGTTTCGATAAGCTTGTATGATTAATACTAGGAGAGC 1584
QY 318 ----- 318
Db 1585 ACATCAATAATTTAGAGTGCAATGACTGATTTAGCCAAATCCTAGCTAGAAATTAATCTG 1644
QY 318 ----- 318
Db 1645 GAAAGAACTTGGAACTCTCAACCATAGGTTTGGTAGAAAATTTGCTTGTGACAGACCA 1704
QY 318 ----- 318
Db 1705 AATGATAGGCTATTGCGCTTGAAATAGTGTTCCTTGTGGTTTCCCAATATTGGAAGTTAAA 1764
QY 318 ----- 318
Db 1765 TCATATGACTTTAGCTGTTGGATACATAATTAAAGCTTAAGCAATGCCAACTCTAAGAAAGTG 1824
QY 319 -----GlyIleValGluSerSerIleLysArgGlyMet 329
Db 1825 TACTTACACAATATCTATTGGTCATAGGTATAGTTGAATCAAGTATCAAGCGTGAATG 1884
QY 330 HisIleGluAlaLeuGluMetValTyrThrPheGlyMetGluAspLysPheSerAlaAla 349
Db 1885 CATATTGAAGCTCTTGAGATGGTTTATACCTTTGGCATGGAGGATAAGTTTTCAGCTGCT 1944
QY 350 LeuValLeuThrSerPheLeuLysMetSerLysGluSerPheGluArgAlaLysArgLys 369
Db 1945 CTAGTTCTTAACCTTCATTTCTTAAAGATGAGCAGAGGTCACTTTGAGAGGGGCAAAACGAAA 2004
QY 370 AlaGlnSerProLeuAlaPheLys----- 377
Db 2005 GCCCAGTCACCGCTGGCATTTGT-ATGAACCTTCCCTTGCACATTTATGTACCTTTATGA 2063
QY 378 -----GluAla 379
Db 2064 ACTCTTTATCATCATCTGAGTCTTGACCATTTGATATATTTATTTCTCAACAGAAAGAGCG 2123
QY 380 AlaThrLysGlnLeuAlaValLeuSerSerValMetGlnCysMetGluThrHisLysLeu 399
Db 2124 GCTACAAAGCAGCTAGCTGTGTTATCATCAGTTATCGAGTGTATGAGACTCACAAGTTA 2183
QY 400 AspProAlaLysGluLeuProGlyTyrGlnIleLysGluGlnIleValSerLeuGluLys 419
```

221	LeuGluSerPheLeuLeuMetProAspArgGlyLysGlyValLysIleGluSerTTP	240
1054	CTGGAGTCCTTTCTCTAAATGCCGTGATCGTGTGAAGGAAAGGTGAAGATTGAGAGTTGG	1113
241	IleLysAspGluAlaGluThrAlaValAlaTTPArgLysArgLeuMetThrGluGly	260
1114	ATTTAAGATGAGCGGAGACGGCTGCTGTCTTGGAGGAAAAGTTGATGACTGAAGA	1173
261	GlyLeuAlaAlaGluLysMetAspAlaArgGlyLeuLeuLeuValAlaCysPhe	280
1174	GGATTAGCTGGCTGAGAAAAAGGATGCAAGGGTTTGCCTTTACTAGTTGCTGTGTTT	1233
281	GlyValProSerIlePheArgSerThrAspLeuLeuAspLeuIleArgMetSerGlySer	300
1234	GGTGTTCCTCTCAAACTTTTGGAGATCAGATTTCCTGGATTGATAAGGATGAGTTCG	1293
301	AsnGluIleAlaGlyAlaLeuLysArgSerGlnPheLeuValProMetValSer	318
1294	AATGAGATTGCCGGTGCTTTGAAGCGGTACAGTTTCTGTCCCTATGGTCTCAGGTACC	1353
318	-----	318
1354	ATATTCTGTTCTCACTCGGTGAATTTCAATGCAAGGTGGTTCCTTTGTTGACATCATC	1413
318	-----	318
1414	GACCAACATCAAGTTCACATCTTTGTTTTCGATAAGCTTGATGGTATAAATAGGAGAC	1473
318	-----	318
1474	ACATCAAAATATTTAGAGTCAATGACTGATGAGCCAAATCCTAGCTAGATAAATTAATCTG	1533
318	-----	318
1534	GAAGAAGCTTGGAACTCTCAACCATAGGTTTGGTACGAATTTGTTGCTTGTACAGAACCA	1593
318	-----	318
1594	AATGATAGGCTATTGCCCTTGAAATAGTGTTCCTTGTGGTTTCCAAATATTGGAAGTTAAAA	1653
318	-----	318
1654	TCATATGACTAGCTGTGGTACTAATTAAGCTTAAGCAATGCCAACTTAAGAAGTGG	1713
319	-----GlyIleValGluSerSerIleLysArgGlyMet	329
1714	TACTTACACAATATTCTATTGGTCATAGGTATAGTTGAATCAAGTATCAAGCGTGAATG	1773
330	HisIleGluAlaLeuGluMetValTyrThrPheGlyMetGluAspLysPheSerAlaAla	349
1774	CATATTGAAGCTCTTGATGGTTTATACCTTTGGCATGGAGGATAAGTTTTCAGCTGCT	1833
350	LeuValLeuThrSerPheLeuLysMetSerLysGluSerPheGluArgAlaLysArgLys	369
1834	CTAGTTCTTAACCTTCTTAAGATGAGCAAGAGAGTCTATTGAGAGGCCAAACGGAAA	1893
370	AlaGlnSerProLeuAlaPheLys	377
1894	GCCAGTCACCGCTGGCATTTTGT-ATGAACCCCTTCCCTTGCACATTATGTACCTTTATGA	1952
378	-----GluAla	379
1953	ACTCTTTATCATCTCGAGTCTGACCATTTGATATATTTATTCTCAACAGAAAGACGG	2012
380	AlaThrLysGlnLeuAlaValLeuSerSerValMetGlnCysMetGluThrHisLysLeu	399
2013	GCTACAAAGCAGTAGCTGTGTATCATCATAGTTATGCAGTGTATGAGACTCACAAGTTA	2072
400	AspProAlaLysGluLeuProGlyTTPGlnIleLysGluClnIleValSerLeuGluLys	419
2073	GATPCTTGCAAAAGAACTACCAAGATGGCAGATCAAGAGCAAAATTGTTAGCTTGGAGAA	2132
420	AspThrLeuGlnLeuAspLysGluMetGluGluLysAlaArgSerLeuSerLeuMetGlu	439

Db	2133	GACACTCTTCAGCTCGCACAAGAGAGATGGAAGAGAAAGCAAGATCTCTCAGTTTAATGGAG	2192
Qy	440	GlulalalaleuhalalysargMettyrAaGlnGlnlleLysArgProAtrLeuSerPro	459
Db	2193	GAAGCCGCATCTCCCAAGAGATGATATAACCAACAGATAAACGTCCTCAAGTGTGTACACC	2252
Qy	460	MetGluMetProProValThrSerSerSerTyrSerProilleTyrArgAspArgSerPhe	479
Db	2253	ATGGAATGCGACACAGTAACCTCTTCATCGTATTCTCCTATCTACCGTGATAGAAGCTTT	2312
Qy	480	ProSerGlnArgAspAspGlnAaPGLulleSerAlaLeuValSerSerTyrLeuGly	499
Db	2313	CCTAGTCAAGAGACAGATGACCAAGATGAAATATCAGCTCTTGAGTAGTTACTCTGGC	2372
Qy	500	ProSerThrSerPheProHisArgSerArgSerProGluTyrMetValProLeuPro	519
Db	2373	CCGTCAACATCTTTCTCATCGCTCAAGAGATCCCGGAATATATGGTTCCACTTCCA	2432
Qy	520	HisGlyGlyLeuGlyArgSerValTyrAlaTyrGluHisLeuAlaProAsnSerTyrSer	539
Db	2433	CATGGTGGTTAGGAAGAGTGTATATGATATGAACATCTGGCCCCAAATTCATCTCT	2492
Qy	540	ProGlyHisGlyHisArgLeuHisArgGlnTyrSerProSerLeuValHisGlyGlnArg	559
Db	2493	CCAGTCAAGGACATAGACTTCATCGACAGTACTCTCCGCTCTTTGGTTACGGACAGAGA	2552
Qy	560	HisProLeuGlnTyrSerProProLleHisGlyGlnGlnLeuProTyrGlyIleGln	579
Db	2553	CATCCACTACAGTACTCTCTCCAAATTCATGACACACACAGTTACCATATGGTATACAA	2612
Qy	580	ArgValTyrArgHisSerProSerGluArgTyrLeuGlyLeuSerAsnGlnArgSer	599
Db	2613	AGGGTTTACACACATTCACCATCTGAAGAAAGATATTGGGTTTATCCAATCAAAGTCT	2672
Qy	600	ProArgSerAsnSerSerLeuAspProLys	609
Db	2673	CCTCGCAGTAACCTCATCTAGACCCCAAA	2702
RESULT 6			
LOCUS	AY198402	3261 bp	DNA linear PLN 12-JUN-2003
DEFINITION	Arabidopsis thaliana ecotype KZ-9 FRIGIDA (FRI) gene, complete cds		
ACCESSION	AY198402		
VERSION	AY198402.1	GI:31558912	
KEYWORDS	Arabidopsis thaliana (thale cress)		
SOURCE	Arabidopsis thaliana		
ORGANISM	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi		
REFERENCE	1 (bases 1 to 3261)		
AUTHORS	Gazzani, S., Gendall, A.R., Lister, C. and Dean, C.		
TITLE	Analysis of the molecular basis of flowering time variation in Arabidopsis accessions		
JOURNAL	Plant Physiol. 132 (2), 1107-1114 (2003)		
REFERENCE	2 (bases 1 to 3261)		
AUTHORS	Gazzani, S. and Dean, C.		
TITLE	Direct Submission		
JOURNAL	Submitted (17-DEC-2002) Cell and Developmental Biology, John Innes Center, Colney Lane, Norwich NR4 7UH, UK		
FEATURES	Location/Qualifiers		
source	1..3261		
	/organism="Arabidopsis thaliana"		
	/mol_type="genomic DNA"		
	/db_xref="taxon:3702"		
	/chromosome="4"		
	/map="between mil1 and mil22"		
	/note="ecotype: KZ-9"		
gene	505..52816		
	/gene="FRI"		
mRNA	Join(<505..1459,1853..2025,2115..52816)		
	/gene="FRI"		

gene
mRNA


```

Db 2373 GAAGCGGACCTTCCAGAGAGATGTATACACACAGATTAACACGTCACAGGTTGTACACC 2432
QY 460 MetGluMetProValThrSerSerTyrrSerProIleTyrrArgAspArgSerPhe 479
Db 2433 ATGGAAATGCCACACAGTAACTCTTCACTGATTTCTCTATCTACCGTGATAGAAGCTTT 2492
QY 480 ProSerGlnArgAspAspGlnAspGluIleSerAlaLeuValSerSerTyrrLeuGly 499
Db 2493 CTTAGTCAAGAGAGATGATACCAAGATGAATATACAGCTCTTGAGTAGTAGTACCTCGGC 2552
QY 500 ProSerThrSerPheProHisArgSerArgArgSerProGluTyrrMetValProLeuPro 519
Db 2553 CCGTCAACATCTTTCTCATCGCTCAAGAAGATCCCGGAATATATGGTTTCCACTTCCA 2612
QY 520 HisGlyGlyLeuGlyArgSerValTyrrAlaTyrrGluHisLeuAlaProAsnSerTyrrSer 539
Db 2613 CATGGTGGGTTAGGAGAGATGTATGATGATGATGATGATGATGATGATGATGATGATGAT 2672
QY 540 ProGlyHisGlyHisArgLeuHisArgGlnTyrrSerProSerLeuValHisGlyGlnArg 559
Db 2673 CCAGGTCACGGACATAGACTTTCATGACAGTACTCTCGTCTTTGGTTACGGACAGAGA 2732
QY 560 HisProLeuGlnTyrrSerProIleHisGlyGlnGlnGlnLeuProTyrrGlyIleGln 579
Db 2733 CATCCACTACACTCTCTCTCCATTCATGACACCAACAGTATACCATATGGTATACAA 2792
QY 580 ArgValTyrrArgHisSerProSerGluGluArgTyrrLeuGlyLeuSerAsnGlnArgSer 599
Db 2793 AGGTTTACAGACATTCACCACTGACAGAGATATTTGGTTTATTCACCAATCAAGTCT 2852
QY 600 ProArgSerAsnSerSerLeuAspProIys 609
Db 2853 CCTCGCAGTAACATCATCATTAGACCCCAA 2882

RESULT 5
AY198403
LOCUS
DEFINITION
Arabidopsis thaliana ecotype Kondara FRIGIDA (FRI) gene, complete cds.
ACCESSION
AY198403
VERSION
AY198403.1 GI:31558914
KEYWORDS
Arabidopsis thaliana (thale cress)
SOURCE
Arabidopsis thaliana
ORGANISM
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopses.
REFERENCE
1 (bases 1 to 3150)
Gazzani S., Gendall A.R., Lister, C. and Dean, C.
Analysis of the molecular basis of flowering time variation in
Arabidopsis accessions
Plant Physiol. 132 (2), 1107-1114 (2003)
REFERENCE
2 (bases 1 to 3150)
Gazzani, S. and Dean, C.
Direct Submission
Submitted (17-DEC-2002) Cell and Developmental Biology, John Innes
Center, Colney Lane, Norwich NR4 7UH, UK
FEATURES
Location/Qualifiers
1. 3150
/organism="Arabidopsis thaliana"
/mol.type="genomic DNA"
/db xref="taxon:3702"
/chromosomes="4"
/map="between ml51 and ml122"
/note="ecotype: Kondara"
<394..>2705
/gene="FRI"
Join(<394..1348,1742..1914,2004..>2705)
/gene="FRI"
/product="FRIGIDA"
Join(394..1348,1742..1914,2004..2705)
/gene="FRI"
CDS

```

```

/note="flowering gene"
/codon_start=1
/product="FRIGIDA"
/protein_id="AAP49809.1"
/db_xref="GI:31558915"
translation="MSNYPTVAQAQPTTANPLLQHQSEQRRELPKIVETESTMD
ITIQSQPOLIKSIDELAAFSVAFTRKQFDLQKHIESIENADLSKESNGVIA
ARNNFQPMSPNNVSVETTVTSQPSQEIPTSNKPEGRNCELMSKGLRKY
IYANISQAKLMBEIPSAKLAKPAKFDVDCIKFYLGRRRAFTKSPMSARQVSL
LILSFLLMDRGKGVKIESWKDBAETAAVAWKRLMTEGGIAAAEKMDARGLILL
VACGVSPNSTDLLDLBMSGNSETAGALKESQFLVPMVSGIVESSIKRGWHIEAL
EMVTFQWNEKPSAALVLTSLFKMSKESPERAKKAQSPKAPKATKOLAVLSVWQ
CMETHKLDPAKELPGWQIKQIVLSERKDLQDKEMEKARSLSLMEERAAKRMVQ
IKKPRISPMEMPVTSVPIYRDRSPFSQDDQDEISALVSYLGFSTSPFPHRS
RRSEYVNPPLPHGGGLGRSVYAEHLAPNSPQGHRLHROVSPSLVHGOQRPLOVSP
PIHQOQLPYGIQVRVHSPSEERYLGLSNQSPRSNSLDPK"
BASE COUNT 951 a 516 c 645 g 938 t
ORIGIN
Alignment Scores:
Pred. No.: 1,27e-186 Length: 3150
Score: 2975.00 Matches: 607
Percent Similarity: 78.83% Conservative: 0
Best Local Similarity: 78.83% Mismatches: 2
Query Match: 95.84% Indels: 162
DB: 8 Gaps: 2
US-09-890-475-1 (1-609) x AY198403 (1-3150)
QY 1 MetSerAsnTyrrProThrValAlaAlaGlnProThrThrAlaAsnProLeuLeu 20
Db 394 ATGTCATATTATCACCGACGTCGCGCGCAACCCACACGACGCGGAATCCACTGCTG 453
QY 21 GlnArgHisGlnSerGluGlnArgArgGluLeuProIysIleValGluThrGluSer 40
Db 454 CAGCGACATCAATCTGAACAGCAGCAAGAAATACCGAAGATTGTGGAACAGAGTCT 513
QY 41 ThrSerMetAspIleThrIleGlyGlnSerIysGlnProGlnPheLeuLysSerIleAsp 60
Db 514 ACAATATGGACATTTACGATCGGTCAATCTNAGCAGCTCAATTTTGAATCCATAGAC 573
QY 61 GluLeuAlaAlaPheSerValAlaValGluThrPheLysArgGlnPheAspLeuGln 80
Db 574 GAATTAGCTCGCTTTTCAGTTGCGAGTGAAACATTTCAACGCCAATTCGATGATCTTCAG 633
QY 81 LysHisIleGluSerIleGluAsnAlaIleAspSerLysLeuGluSerAsnGlyValVal 100
Db 634 AGGCATCTGAGTCAATCGAAACGCAATTTGATTCGAACCTCGAGAGTAACGGCTTGTCTC 693
QY 101 LeuAlaAlaArgAsnAsnAsnPheHisGlnProMetLeuSerProProArgAsnVal 120
Db 694 CTCGCGCGCGGCAACAATAATTTCCATCAGCCGATGTTATCGCTCGCGCAACAATGTA 753
QY 121 SerValGluThrThrValThrValSerGlnProSerGlnGluIleValProGluThrSer 140
Db 754 TCTGTAGAAACACCGCTCAGTGTAGCCACCGCTCTCAGGAGATTGTACCGAGACGCTG 813
QY 141 AsnLysProGluGlyGlyArgMetCysGluLeuMetCysSerLysGlyLeuArgLysTyr 160
Db 814 AATAAACCGGAGCGGAGCGATGTTGTCAGTTGATGTGTACAAAGGTCTGCGTAATATC 873
QY 161 IleTyrrAlaAsnIleSerAspGlnAlaLysLeuMetGluGluIleProSerAlaLeuLys 180
Db 874 ATATACGCAATATCTCTGATCAAGCTAAGTTAATGGAAGAGATTCCTTCCAGCTTTGAAA 933
QY 181 LeuAlaLysGluProAlaLysPheValLeuAspCysIleGlyLysPheTyrrLeuGlnGly 200
Db 934 TTGCGCAAGAGACCGAGGAGTTTGTATTTGATTTGATTTGCAAGTTTACTTACAGGG 993
QY 201 ArgArgAlaPheThrLysGluSerProMetSerSerAlaArgGlnValSerLeuLeulle 220
Db 994 CGTAGACATTTTACTAAGAGTCGCTATGAGCTCTGCGAGACAAAGTTTCGCTTCTTATA 1053

```

```

/db_xref="GI:10801174"
/translation="WSNYPPTVAQPTTANPILLOPHCSORRELPKIVENESMSD
ITIGSKQDQFLKSIDELAAARSVAVERFKQFDLDLQKHIESINADISKLENGVLA
ARNNNFHPMLSPNNVSVETTVVSPQSGEIVFSTNSPEGRKCELMCKSLGRKY
IYANTSDQAKMEBIPSKALUKAEKFAKFDVDCIGFKYLOQRAPFTKESPMSSARQVSL
LILESFLMPDRGKVKIWEIKDEAETAARVAKELMTEGGALAAEKMDARGLLIL
VACFVPSNFSATDILLDIRMSGNEIAGALKRSQFLVPMVSGIIVESSIKRGMHIEAL
EMVYTFGMEDKFSALVLVTSFLKMSKESPERAKRKAQSPKAFKAATKOLAVLSVWQ
CWEHLDPKXELPMQWIKEOIVSLEKDTLQDKEMEKEARSLSLMEEAALAKRMVNO
QIKRPLSPMBMPVWTSSVSPYVDRSPSDDODEISALVSVLSPGSPSPFPRS
RASPEYVPLPHGGLGRSVAYEHLAPNSVSPGHGHLRHOYFSPSLVHGQRHPLQSP
PIHQQQQLPYGIQVRYRHSPEERLYLGNQSPRSNSLDPK"
BASE COUNT 1174 a 717 c 775 g 1095 t
ORIGIN

```

```

Alignment Scores:
Pred. No.: 6 43e-187 Length: 3761
Score: 2981.00 Matches: 608
Percent Similarity: 79.96% Conservative: 0
Best Local Similarity: 79.96% Mismatches: 1
Query Match: 96.04% Indels: 162
DB: 8 Gaps: 2

```

US-09-890-475-1 (1-609) x AF228499 (1-3761)

```

QY 1 MetSerAsnTyrProProThrValAlaAlaGlnProThrThrAlaAsnProLeuLeu 20
DB 574 ATGTCCAAATATCCACCGAGCGTGGCGCGCAACCCACACGACGCGGAATCCACTGCTG 633
QY 21 GlnArgHisGlnSerGluGlnArgArgGluLeuProLysIleValGluThrGluSer 40
DB 634 CAGCGACATCAATCTGAACAGCGAGAGAGAAATACCGAGATTGTGGAACAGAGTCT 693
QY 41 ThrSerMetAspIleThrIleGlyGlnSerLysGlnProGlnPheLeuLysSerIleAsp 60
DB 694 ACAAGTATGCATACGATCGGTCAATCTAAGCAGCGCTCAATTTTGAAATCCATAGAC 753
QY 61 GluLeuAlaAlaPheSerValAlaValGluThrPheLysArgGlnPheAspAspLeuGln 80
DB 754 GAATTAGCTGCGTTTTTCAGTTCAGTGGAAACATTCAAACCCCAATTCGATGATCTTCAG 813
QY 81 LysHisIleGluSerIleGluAsnAlaIleAspSerLysLeuGluSerAsnArgIleValVal 100
DB 814 AAGCACATCGATCAATCGAAACCAATTCATTCCAAACTCGAGAGTAAACGGCGTGTGTC 873
QY 101 LeuAlaAlaArgAsnAsnAsnPheHisGlnProMetIleuSerProArgAsnAsnVal 120
DB 874 CTCGCCGCGGGAACAATAATTTCCATCAGCCGATGTTATCGCTCCGCGGAACAATGTA 933
QY 121 SerValGluThrThrValThrValSerGlnProSerGlnGluIleValProGluThrSer 140
DB 934 TCTGTAGAAACCCACCGTCACTGTGAGCCAAACCGTCTCAGGAGATTGTACCGAGACGTCG 993
QY 141 AsnLysProGluGlyGlyArgMetCysGluLeuMetCysSerLysGlyLeuArgLysTyr 160
DB 994 AATAAACCGAGGGGGGAGCGTATGTGTGAGTTGATGTAGTACAAAGGTCTCGCTAAATAC 1053
QY 161 IleTyrAlaAsnIleSerAspGlnAlaLysLeuMetGluGluIleProSerAlaLeuLys 180
DB 1054 ATATACCGGAATATCTCTGATCAAGCTAAGTTAATGGAGAGATTCCTTCAGCTTGAAA 1113
QY 181 LeuAlaLysGluProAlaLysPheValLeuAspCysIleGlyLysPheTyrLeuGlnGly 200
DB 1114 TTGGCCAGAGGACGCGAAGTTTGTATTTGATTGTATTTGGCAAGTTTATCTTACAAGGG 1173
QY 201 ArgArgAlaPheThrLysGluSerProMetSerSerAlaArgGlnValSerLeuLeuIle 220
DB 1174 CGTAGAGCATTTACTTAAGAGTCGCGCTATGAGCTCTCGAGACCAAGTTTCGCTTCTATA 1233
QY 221 LeuGluSerPheLeuLeuMetProAspArgGlyLysGlyLysValLysIleGluSerTrp 240
DB 1234 CTGGAGTCTTTCTTCTAATGCTGATCGTGTAAAGGGAAGGTCAAGATTGAGAGTTGG 1293

```

```

QY 241 IleLysAspGluAlaGluThrAlaAlaValAlaTrpArgLysArgLeuMetThrGluGly 260
DB 1294 ATTTAAAGATGAGCGGAGACGGCTGCTGTTGCTTGAGGAAAGGTTGATGACTGAAGA 1353
QY 261 GlyLeuAlaAlaAlaGluLysMetAspAlaArgGlyLeuLeuLeuValAlaCysPhe 280
DB 1354 GGATTAGCTCGCGCTGAGAAATTCGATGCAAGGGGTTTCTTTTACTAGTTCGCTGTTT 1413
QY 281 GlyValProSerAsnPheArgSerThrAspLeuLeuAspLeuIleArgMetSerGlySer 300
DB 1414 GGTGTTCTTCAACTTTAGAGTACAGATTGCTGGAATTTGATAAGGATGAGTGGTTCG 1473
QY 301 AsnGluIleAlaGlyAlaLeuLysArgSerGlnPheLeuValProMetValSer----- 318
DB 1474 AATGAGATTGCGGTGCTTTTCAAGCGGTACAGTTTCTTTGCTCCCTATGCTCAGGTACC 1533
QY 318 ----- 318
DB 1534 ATATTCTGTTCTACTCGGTGAATTTCAITGCAAAAGTGGTTCCTTTTGTGACATCATC 1593
QY 318 ----- 318
DB 1594 GACCAACATCAAGTTCCATCTTGTGTTTTCGATAAGCTTGATGGTATAAACTAGGAGAGC 1653
QY 318 ----- 318
DB 1654 ACATCAAAATATTAGATGCAATGACTGATTGAGCCAAATCCCTAGTAGAATTAATCTG 1713
QY 318 ----- 318
DB 1714 GAAAGAACTTGGAACTCTCAACCATAGTTTGGTACGAAATTTGTTGCTTGTGAGAACCA 1773
QY 318 ----- 318
DB 1774 AATGATAGGCTATTGCGCTTGAAATAGTGTTCCTGTTGTTTCCAAATATTGGAAGTTAAA 1833
QY 318 ----- 318
DB 1834 TCGTATGATTAGCTGTTGGATACATAATTAAGCTTAAGCAATCCCAACTCTAAGAAGTGG 1893
QY 319 ----- GlyIleValGluSerSerIleLysArgGlyMet 329
DB 1894 TACTTACACAAATATTCTTATTGGTTCATAGGTATAGTTGAATCAAGTATCAAGCGTGAAG 1953
QY 330 HisIleGluAlaLeuGluMetValTyrThrPheGlyMetGluAspLysPheSerAlaAla 349
DB 1954 CATATTGAGCTCTTGAGATGGTTTATACCTTTGGCATGGAGGATAAGTTTTCAGTGTCT 2013
QY 350 LeuValLeuThrSerPheLeuLysMetSerLysGluSerPheGluArgAlaLysArgLys 369
DB 2014 CTAGTTCTTAATCTTCAATCTTAAAGATGAGCAAGGAGTCAATTTGAGAGGGCAAAACGAAA 2073
QY 370 AlaGlnSerProLeuAlaPheLys----- 377
DB 2074 GCCCAGTCCACCGCTGGCATTTGT-ATGAACCCCTCCCTTGCACATTATGTACCTTTATGA 2132
QY 378 ----- GluAla 379
DB 2133 ACTCTTTATCATCTCTGAGTCTGACCATTTGATATATTTTCTCAACAGAAAGAGCG 2192
QY 380 AlaThrLysGlnLeuAlaValIleuSerSerValMetGlnCysMetGluThrHisLysLeu 399
DB 2193 GCTACAAAGCAGCTAGCTGTGTTATCATCATGTTATGAGTGTATGAGAGACTCAAGTTA 2252
QY 400 AspProAlaLysGluLeuProGlyTrpGlnIleLysGluGlnIleValSerLeuGluLys 419
DB 2253 GATCTCGGAAAGAACTACCAGGATGCGAGATCAAGAGCAAAATTTGTAGCTTGGAGAAA 2312
QY 420 AspThrLeuGlnLeuAspLysGluMetGluGluLysAlaArgSerLeuSerLeuMetGlu 439
DB 2313 GACACTCTTCAGCTCGCAAGAGATGGAAGAGAGATCTCTCAGTTTAAATGGAG 2372
QY 440 GluAlaAlaLeuAlaLysArgMetTyrAsnGlnIleLysArgProArgLeuSerPro 459

```

1354 GGATTAGCTGGCTGAGAAATGGATGATCGAGGGTTTGTCTTTTACTAGTTGCTTGTGTTT 1413
QY GlyValProSerAsnPhaArgSerThrAspLeuLeuAspLeuLeuIleArgMetSerGlySer 300
Db 1414 GGTGTTCCTTCAAACTTTAGAGTACAGATTGTCTGGATTGTGATTAAGATGAGTGTTCG 1473
QY 301 AsnGluIleAlaGlyAlaLeuIleArgSerGlnPheLeuValProMetValSer----- 318
Db 1474 AATGAGATTGCGGTGCTTTGAGGGGTACAGTTTCTTGTCCCTATGGTCTCAGGTACC 1533
QY 318 ----- 318
Db 1534 ATATTCTGTCTCTACTCGGTGAATTCATTGCAAGGTGTTCCCTTTGTGTGATCATC 1593
QY 318 ----- 318
Db 1594 GACCAACATCAAGTCCATCTTTGTTTTTCGATAAGCTTGTGATATAACTAGGAGGC 1653
QY 318 ----- 318
Db 1654 ACATCAATATTAGAGTCAATGACTGATTGAGCCAAATCCTAGCTAGAAATTAATCTG 1713
QY 318 ----- 318
Db 1714 GAAAGAACTTGAACCTCAACCATAGTGTGTTGTTGTCAGAAATTGCTGTCTCAGAACCA 1773
QY 318 ----- 318
Db 1774 AATGATAGCTATTGCTTGAATAGTGTCTTGTGTTGTTCCCAATATTGGAAGTTAAAA 1833
QY 318 ----- 318
Db 1834 TCGTATGACTTAGCTGTGGATACATAAAGCTTAAGCAATGCAACTCTAAGAAGTGG 1893
QY 319 -----GlyIleValGluSerSerIleLysArgGlyMet 323
Db 1894 TACTTACAAATATTCTATTGCTCATAGTATGATTGAATCAAGTATCAAGCGTGGATG 1953
QY 330 HisIleGluAlaLeuMetValThrPheGlyMetGluAspLysPheSerAlaAla 349
Db 1954 CATATTGAAGCTCTTGAGATGGTTTATACCTTTGGCATGGAGGATAAGTTTTCAGCTGCT 2013
QY 350 LeuValLeuThrSerPheLeuLysMetSerLysGluSerPheGluAlaLysArgLys 369
Db 2014 CTAGTTCTAACCTTCTTAAAGATGACGAGGAGTCACTTTGAGGGCCAAACGGA 2073
QY 370 AlaGlnSerProLeuAlaPheLys----- 377
Db 2074 GCCAGTCAACGCTGGCATTTGT-ATGAACCTTCCCTTGCACTATTATGTACCTTTATGA 2132
QY 378 -----GluAla 379
Db 2133 ACTCTTTATCATCTGAGTCTGACCTGACCTGATGATATATTATTCTCAACAGAAAGACG 2192
QY 380 AlaThrLysGlnLeuAlaValLeuSerSerValMetGlnCysMetGluThrHisLysLeu 399
Db 2193 GCTTACAAAGCAGTAGCTGTGTTATCATCAGTTATGCTAGTGTATGGAGACTCACAAGTTA 2252
QY 400 AspProAlaLysGluLeuProGlyTrpGlnIleLysGluGlnIleValSerLeuGluLys 419
Db 2253 GATCTTGGGAAAGAACTACAGAGTGGCAGATCAAGAGCAAAATGTTAGCTTGGAGAAA 2312
QY 420 AspThrLeuGlnLeuAspLysGluMetGluGluLysAlaArgSerLeuSerLeuMetGlu 439
Db 2313 GACACTCTTCAGCTCGACAAAGAGATGGAAGAGAAAGCAAGATCTCTCAGTTTAATGGAG 2372
QY 440 GluAlaAlaLeuAlaLysArgMetThrAsnGlnGlnIleLysArgProArgLeuSerPro 459
Db 2373 GAAGCCGCACTTGGCCAAAGAAATGATATACCAACAGATATAACCGTCCAAAGTTGTCCACC 2432
QY 460 MetGluMetProProValThrSerSerSerThrSerProIleThrArgAspArgSerPhe 479
Db 2433 ATGGAATATGCCACCAAGTAATCTTCTTCATCGTATTCTCTATCTACCGTGATAGAAGCTTT 2492

QY 480 ProSerGlnArgAspAspGlnAspGluIleSerAlaLeuValSerSerThrLeuGly 499
Db 2493 CTTAGTCAAGAGACGATGACCAAGATGAATATCAGCTCTTGTGAGTAGTTACCTCGGC 2552
QY 500 ProSerThrSerPheProHisArgSerArgArgSerProGluThrMetValProLeuPro 519
Db 2553 CCGTCAACATCTTTCTCTCATCTCAAGAAATCCCGGAATATATGTTCCACTTCCA 2612
QY 520 HisGlyGlyLeuGlyArgSerValTyrAlaTyrGluHisLeuAlaProAsnSerThrSer 539
Db 2613 CATGGTGGTTAGGAAGAGTGTATATGCATATGAACATCTGGCCCCCAATCATACTCT 2672
QY 540 ProGlyHisGlyHisArgLeuHisArgGlnTyrSerProSerLeuValHisGlyGlnArg 559
Db 2673 CCAGTCCAGGACATAGACTTCATCGACAGTACTCTCCGCTTTGGTTCACGACAGAGA 2732
QY 560 HisProLeuGlnTyrSerProProIleHisGlyGlnGlnGlnLeuProTyrGlyIleGln 579
Db 2733 CATCCACTACAGTACTCTCTCCAAATTCATGGACACACAGTTACCATATGGTATACAA 2792
QY 580 ArgValTyrArgHisSerProSerGluGluArgTyrIleuGlyLeuSerAsnGlnArgSer 599
Db 2793 AGGTTTACAGACATTCCCATCTGAAGAAAGATATTGGGTTTATCCAAATCAAGGTCT 2852
QY 600 ProArgSerAsnSerSerLeuAspProLys 609
Db 2853 CCTCGCAGTAATCATCTAGACCCCAA 2882
RESULT 4
AF228499 3761 bp DNA linear PLN 14-OCT-2000
LOCUS Arabidopsis thaliana FRIGIDA (FRI) gene, complete cds.
DEFINITION Arabidopsis thaliana FRIGIDA (FRI) gene, complete cds.
ACCESSION AF228499
VERSION AF228499.1 GI:10801173
KEYWORDS
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
REFERENCE 1 (bases 1 to 3761)
AUTHORS Johanson,U., West,J., Lister,C., Michaels,S., Amasino,R. and Dean,C.
TITLE Molecular analysis of FRIGIDA, a major determinant of natural variation in Arabidopsis flowering time
JOURNAL Science 290 (5490), 344-347 (2000)
MEDLINE 20485641
PUBMED 11030654
REFERENCE 2 (bases 1 to 3761)
AUTHORS Johanson,U. and Lister,C.
TITLE Direct Submission
JOURNAL Submitted (25-JAN-2000) Molecular Genetics, John Innes Centre, Colney Lane, Norwich, Norfolk NR4 7UH, U.K.
FEATURES
Location/Qualifiers
1..3761
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/cultivar="ecotype H51"
/db_xref="taxon:3702"
/chromosome="4"
/map="between m151 and m122"
/574..>2885
/gene="FRI"
/gene="FRI"
/product="FRIGIDA"
join(574..1528,1922..2094,2184..>2885)
/gene="FRI"
/product="FRIGIDA"
join(1574..1528,1922..2094,2184..2885)
/note="late flowering gene"
/codon_start=1
/product="FRIGIDA"
/protein_id="AAG23414.1"

341 GlyMetGluAspLysPheSerAlaAlaLeuValLeuThrSerPheLeuLysMetSerLys 360
 Db GGATGGAGGATAGTTTTCAGCTGCTAGTCTTAACCTTATTCATTCTTAAGATGAGCAAG 1441
 361 GluSerPheGluArgAlaLysArgLysAlaGlnSerProLeuAlaPheLysGluAlaAla 380
 Db GAGTCATTTGAGAGGGCAAAAGCAAGCCCGAGTCCCGCTGGCATTTTAAAGAGCGGCT 1501
 381 ThrLysGlnLeuAlaValLeuSerSerValMetGlnCysMetGluThrHisLysLeuAsp 400
 Db ACAAAGCAGCTAGCTGTATCATCAGTTATGCGAGTGTATGAGACTCACAGTTAGAT 1561
 401 ProAlaLysGluLeuProGlyTrpGlnLysGluGlnLeuValSerLeuGluLysAsp 420
 Db CTTGCGAAGAACTACCAAGATGCCAGATCAAGAGCAAAATTTGTTAGCTTGAGAGAAC 1621
 421 ThrLeuGlnLeuAspLysGluMetGluGluLysAlaAlaArgSerLeuSerLeuMetGluGlu 440
 Db ACTCTTCAGCTCGACAAGAGATGGAGAGAGAAAGCAAGATCTCTCAGTTTATGGAGAA 1681
 441 AlaAlaLeuAlaLysArgMetTyrAsnGlnGlnLysLeuArgProArgLeuSerProMet 460
 Db GCCGCACTTCCCAAGAGATGTTATACCAACAGATAAAAGCTCCAAAGTTGTACCCCATG 1741
 461 GluMetProValThrSerSerSerTyrSerProLysTyrArgAspArgSerPhePro 480
 Db GAAATGCCACGTAACCTCTTCATCGTATCTCCCTATCTACCGTATAGAAAGCTTCTCT 1801
 481 SerGlnArgAspAspGlnAspGluLysLeuSerAlaLeuValSerSerTyrLeuGlyPro 500
 Db AGTCAAGAGACGATGACCAAGATGAATATCAGCTCTTGTGAGTAGTTACCTCGGCCCG 1861
 501 SerThrSerPheProHisArgSerArgSerProGluTyrMetValProLeuProHis 520
 Db TCAACATCTTTTCTCATCTCCCTCAAGAGATCCCGGAATATATGTTCCACTTCCACAT 1921
 521 GlyGlyLeuGlyArgSerValTyrAlaTyrGluHisLeuAlaProAsnSerTyrSerPro 540
 Db GTGGGTAGGAAGAGTGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1981
 541 GlyHisGlyHisArgLeuHisArgGlnTyrSerProSerLeuValHisGlyGlnArgHis 560
 Db GGTCAAGGACATAGACTTCATGACAGTACTCTCCGTCTTTGGTTCCAGGACAGAGACAT 2041
 561 ProLeuGlnTyrSerProProLysHisGlyGlnGlnLeuProTyrGlyLysGlnArg 580
 Db CCACTACAGTACTCTCTCCATTCATGCAACACAGTACCATGATGATGATGATGATGATGAT 2101
 581 ValTyrArgHisSerProSerGluGluArgTyrLeuGlyLeuSerAsnGlnArgSerPro 600
 Db GTTTACAGACATTCACCATCTGAAGAAAGATATTGGTTTATCCCAATCAAAAGGTCTCTCT 2161
 601 ArgSerAsnSerSerLeuAspProLys 609
 Db CGCAGTAACCTCATCTAGACCCCAAA 2188
 RESULT 3
 LOCUS AX032762 3761 bp DNA linear PAT 21-SEP-2000
 DEFINITION Sequence 2 from Patent WO0046358.
 ACCESSION AX032762
 VERSION AX032762.1 GI:10279738
 KEYWORDS Arabidopsis thaliana (thale cress)
 SOURCE Arabidopsis thaliana
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi-
 s
 1
 REFERENCE Dean, C., West, J., and Johanson, J.
 AUTHORS Plant Gene
 TITLE Patent: WO 0046358-A 2 10-AUG-2000;
 JOURNAL DEAN CAROLINE (GB); WEST JOANNE (GB); PLANT BIOSCIENCE LTD (GB);

JOHANSON URBAN (SE)
 Location/Qualifiers
 1. 3761
 /organism="Arabidopsis thaliana"
 /mol_type="genomic DNA"
 /db_xref="taxon:3702"
 BASE COUNT 1174 a 717 c 775 g 1095 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 6,43e-187 Length: 3761
 Score: 2981.00 Matches: 608
 Percent Similarity: 78.9% Conservative: 0
 Best Local Similarity: 78.9% Mismatches: 1
 Query Match: 96.04% Indels: 162
 DB: Gaps: 2
 US-09-890-475-1 (1-609) x AX032762 (1-3761)
 QY 1 MetSerAsnTyrProProThrValAlaAlaGlnProThrThrAlaAsnProLeuLeu 20
 Db 574 ATGTCCAAATTTCCACCGAGCGTGGCGGCAACCCACAGCGGCGGAATCCACACTGCTG 633
 QY 21 GlnArgHisGlnSerGluGlnArgArgGluLeuProLysIleValGluThrGluSer 40
 Db 634 CAGCGACATCAATCTGAACAGCGACGAAGAGATTCGCAAGATTGTCGAACACAGAGTCT 693
 QY 41 ThrSerMetAspIleThrIleGlnSerLysGlnProGlnPheLeuLysSerIleAsp 60
 Db 694 ACAAGTATCGACATTACGATCGTCAATCTAAGCAGCCCTCAATTTTGAATCCATAGAC 753
 QY 61 GluLeuAlaAlaPheSerValAlaValGluThrPheLysArgGlnPheAspAspLeuGln 80
 Db 754 GAATTAGCTGGCTTTTCAGTTGCTGAGTGAACACATTCAAACGCCAATTCGATGATCTTCAG 813
 QY 81 LysHisIleGluSerIleGluAsnAlaIleAspSerLysLeuGluSerAsnGlyValVal 100
 Db 814 AAGCACATCGAGTCAATCGAAACCGCAATTCATCCAACTCGAGAGTAACGGGTGTTC 873
 QY 101 LeuAlaAlaArgAsnAsnAsnPheHisGlnProMetLeuSerProProArgAsnAsnVal 120
 Db 874 CTGCGCGCGGGAACATATATTTCCATCAGCGGATGTTATCGCTCCGCGGAACAATGTA 933
 QY 121 SerValGluThrThrValThrValSerGlnProSerGlnGlnIleValProGluThrSer 140
 Db 934 TCTGTAGAAACCACTCTACTGTGAGCAACCGTCTCAGGAGATTGTACCGGACGCTCG 993
 QY 141 AsnLysProGluGlyGlyArgMetCysGluLeuMetCysSerLysGlyLeuArgLysTyr 160
 Db 994 AATAAACCGGAGGGGAGCGTATGTGTAGTTGATGTGTAGCAAGGTCTCGGTAATATAC 1053
 QY 161 IleTyrAlaAsnIleSerAspGlnAlaLysLeuMetGluGluIleProSerAlaLeuLys 180
 Db 1054 ATATACGCAATATCTCTGATCAAGCTAAGTTAATGAAGAGATTCTTCAGCTTTGAAA 1113
 QY 181 LeuAlaLysGluProAlaLysPheValLeuAspCysIleGlyLysPheTyrLeuGlnGly 200
 Db 1114 TTGGCAAGGAGCGGAGGAGTTGTATTGGATTGTATTGGCAAGTTTACTTACAGGG 1173
 QY 201 ArgArgAlaPheThrLysGluSerProMetSerSerAlaAlaArgGlnValSerLeuLeu 220
 Db 1174 CGTAGACATTTACTTAAGAGTCGCTATGAGCTCTCGGAGACAAGTTTCGCTCTTATA 1233
 QY 221 LeuGluSerPheLeuLeuMetProAspArgGlyLysGlyLysValIleGluSerTrp 240
 Db 1234 CTGGAGTCTTTTCTTAATGCTGATCGTGTAAAGGAAGGAGGAAAGATTGAGATTGG 1293
 QY 241 IleLysAspGluAlaGluThrAlaAlaValAlaTyrArgLysArgLeuMetThrGluGly 260
 Db 1294 ATTAAAGATGAGCGGAGACGGCTGCTGTGCTTGAGGAAAGGTTGATGACTCAAGGA 1353
 QY 261 GlyLeuAlaAlaAlaGluLysMetAspAlaAlaArgGlyLeuLeuLeuValAlaCysPhe 280

Db 1201 CCTGCGAAGAACTACCAAGGATGGCAGATCAAGAGCAAAATTGTTAGCTGGGAGAAGAC 1260
QY 421 ThrLeuGlnLeuAspLysGluMetGluGlyLysAlaArgSerLeuSerLeuMetGluGlu 440
Db 1261 ACTCTTCAGCTCGACAAGAGATGGAAGAGAAAGCAAGATCTCTCAGTTTAATGGAGAA 1320
QY 441 AlaAlaLeuAlaLysArgMetTyrAsnGlnGlnLysArgProArgLeuSerProMet 460
Db 1321 GCCGCACCTTGCACAGAGAATGTATACCAACAGATAAACCTCAAGGTTGTCAACCATG 1380
QY 461 GluMetProProValThrSerSerTyrSerProIleTyrArgAspArgSerPhePro 480
Db 1381 GAAATGCCACAGTAACTCTTCATCGTATCTCTCTATCTACCGTGATAGAGCTTTCCT 1440
QY 481 SerGlnArgAspAspGlnAspGluLysSerAlaLeuValSerSerTyrLeuGlyPro 500
Db 1441 AGTCAAGAGAGCGATGACCAAGATGAATATACAGCTCTTGTGAGTAGTACCTCGGCCG 1500
QY 501 SerThrSerPheProHisArgSerArgSerProGluTyrMetValProLeuProHis 520
Db 1501 TCAACATCTTTTCCATCGCTCAAGAGATCCCGGATATATGGTTCCACTTCCACAT 1560
QY 521 GlyGlyLeuGlyArgSerValTyrAlaTyrGluHisLeuAlaProAsnSerTyrSerPro 540
Db 1561 GTGGGTTAGGAAGAAGTGTATATGCATATGAACATCTGGCCCCAAATTCATCTCTCCA 1620
QY 541 GlyHisGlyHisArgLeuHisArgGlnTyrSerProSerLeuValHisGlyGlnArgHis 560
Db 1621 GGTACGGACATAGACTTCATCGACAGTACTCTCCGTCTTTGGTTCCAGGACAGACAT 1680
QY 561 ProLeuGlnTyrSerProProIleHisGlyGlnGlnLeuProTyrGlyLeuArg 580
Db 1681 CCACATACGACTCTCTCTCAATTCATGACAAACACAGTGTACCATATGGTATACAAAG 1740
QY 581 ValTyrArgHisSerProSerGluGlnArgTyrLeuGlyLeuSerAsnGlnArgSerPro 600
Db 1741 GTTTACAGACATTCACCATCTGAAGAAGATATTTGGGTTTATCCAAATCAAGGTCTCCT 1800
QY 601 ArgSerAsnSerSerLeuAspProLys 609
Db 1801 CGCAGTAACTCATATAGACCCCAAA 1827

RESULT 2
AX032763 2257 bp DNA linear PAT 21-SEP-2000
LOCUS Sequence 3 from Patent WO0046358.
DEFINITION AX032763
ACCESSION AX032763
VERSION AX032763.1 GI:110279739
KEYWORDS synthetic construct
SOURCE synthetic construct
ORGANISM artificial sequences.
REFERENCE 1
AUTHORS Dean C., West, J. and Johanson, U.
TITLE Plant gene
JOURNAL Patent: WO 0046358-A 3 10-AUG-2000;
DEAN CAROLINE (GB); WEST JOHANN (GB); PLANT BIOSCIENCE LTD (GB);
JOHANSON URBAN (SE)
FEATURES
source
1..2257
/organism="synthetic construct"
/mol_type="genomic DNA"
/db_xref="taxon:32630"
/note="likely cDNA sequence of the H51 FRI gene"

BASE COUNT 698 a 443 c 493 g 623 t
ORIGIN
Alignment Scores:
Pred. No.: 2 58e-195 Length: 2257
Score: 3104.00 Matches: 609
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0

Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0
US-09-890-475-1 (1-609) x AX032763 (1-2257)
QY 1 MetSerAsnTyrProProThrValAlaAlaGlnProThrThrAlaAsnProLeuLeu 20
Db 362 ATGTCATATTATCCACGCGGTGGCGCGCAACCCACACGCGCGAATCCACTGCTG 421
QY 21 GlnArgHisGlnSerGluGlnArgArgGluLeuProLysIleValGluThrGluSer 40
Db 422 CAGCGACATCAATCTGACACGCGACGAGAGAAATACCGAAGATTGTCCGAAACAGAGTCT 481
QY 41 ThrSerMetAspIleThrIleGlyGlnSerLysGlnProGlnPheLeuLysSerIleAsp 60
Db 482 ACAAGTAGGACATTACGATCGGTCAATCAAGCAGCCTCAATTTTGAATCCATAGAC 541
QY 61 GluLeuAlaAlaPheSerValAlaValGluThrPheLysArgGlnPheAspLeuGln 80
Db 542 GAAATTAGCTGCTTTTTCAGTTTCAGTGGCAACATTCAAACGCCAATTCGATGATCTTCAG 601
QY 81 LysHisIleGluSerIleGluAsnAlaIleAspSerLysLeuGluSerAsnGlyValVal 100
Db 602 AAGCACATCGAGTCAATCGAAGAACGCAATTTGATTCCAACTCGAGAGTAAACGCGGTGTC 661
QY 101 LeuAlaAlaArgAsnAsnAsnPheHisGlnProMetLeuSerProProArgAsnAsnVal 120
Db 662 CTCGCGCGCGGAGCAATAATTTCCATCAGCCGATGTTATCGCCTCCGCGGAACAATGTA 721
QY 121 SerValGluThrThrValThrValSerGlnProSerGlnGluIleValProGluThrSer 140
Db 722 TCTGTAGAAACCCCGCTCATCTGTAGCCACCCGCTCTCAGAGATTGTACCGGAGACGTCG 781
QY 141 AsnLysProGluGlyGlyArgMetCysGluLeuMetCysSerLysGlyLeuArgLysTyr 160
Db 782 AATAAACCGGAGGGGAGCTATGTGAGTTGATGTGTAGCAAGGCTCTGCGTAATATAC 841
QY 161 IleTyrAlaAsnIleSerAspGlnAlaLysLeuMetGluGluIleProSerAlaLeuLys 180
Db 842 ATATACGGCAATATCTCTGATCAAGCTAAGTTAATGGAAGAGATTCTCTCAGCTTTGAAA 901
QY 181 LeuAlaLysGluProAlaLysPheValLeuAspCysIleGlyLysPheTyrLeuGlnGly 200
Db 902 TTGGCCAAAGAGGACGAGCAAGTTTGTATTGATTGTTATTTGGCAAGTTTACTTACAAAGG 961
QY 201 ArgArgAlaPheThrLysGluSerProMetSerSerAlaArgGlnValSerLeuLeuIle 220
Db 962 CGTAGAGCAATTTACTAAAGAGTCGCTATGAGCTCTCGAGACAAGTTTCGCTTCTTATA 1021
QY 221 LeuGluSerPheLeuLeuMetProAspArgGlyLysGlyLysValLysIleGluSerTyr 240
Db 1022 CTGGAGTCTTTTCTTCTAATGCCGTGATCGTGTAAAGGAAAGGTTGAAGATTGAGAGTTGG 1081
QY 241 IleLysAspGluAlaGluThrAlaAlaValAlaTyrArgLysArgLeuMetThrGluGly 260
Db 1082 ATTAAAGATGAGCGGAGACGCTCTCTGCTTGGAGAAAAGGTTGATGACTGAAGGA 1141
QY 261 GlyLeuAlaAlaGluLysMetAspAlaArgGlyLeuLeuLeuValAlaLysPhe 280
Db 1142 GGATTAGCTGCGGCTGAGAAAATGGAATGGAAGGGTTTTCCTTTTACTAGTTGCTTTT 1201
QY 281 GlyValProSerAsnPheArgSerThrAspLeuLeuAspLeuIleArgMetSerGlySer 300
Db 1202 GGTGTTCTCTCAAACTTTAGGAGTACAGATTGCTGGATTTTGTAAAGATGAGTGGTTCG 1261
QY 301 AsnGluIleAlaGlyAlaLeuLysArgSerGlnPheLeuValProMetValSerGlyIle 320
Db 1262 AATGAGATTGCGGCTCTTTGAAGCGGTCACAGTTTCTTGTCCTCTAGGTATATA 1321
QY 321 ValGluSerSerIleLysArgGlyMetHisIleGluAlaLeuGluMetValTyrThrPhe 340
Db 1322 GTTGAATCAAGTATCAAGCGTGGAAATGCATATTTGAAGCTCTTTAGATGGTTTATACCTTT 1381

```

AF228500      1830 bp      mRNA      linear      PLN 14-OCT-2000
LOCUS         Arabidopsis thaliana FRIGIDA (FRI) mRNA, complete cds.
DEFINITION   Arabidopsis thaliana FRIGIDA (FRI) mRNA, complete cds.
ACCESSION   AF228500
VERSION      AF228500.1 GI:10801175
KEYWORDS
SOURCE
ORGANISM     Arabidopsis thaliana (thale cress)
              Arabidopsis thaliana
              Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
              Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
              rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE    1 (bases 1 to 1830)
              Johanson,U., West,J., Lister,C., Michaels,S., Amasino,R. and
              Dean,C.
              Molecular analysis of FRIGIDA, a major determinant of natural
              variation in Arabidopsis flowering time
              Science 290 (5490), 344-347 (2000)
MEDLINE      20485641
FUEMED       11030654
REFERENCE    2 (bases 1 to 1830)
              Johanson,U. and Lister,C.
              Direct Submission
              Submitted (25-JAN-2000) Molecular Genetics, John Innes Centre,
              Colney Lane, Norwich, Norfolk NR4 7UH, U.K.
FEATURES
source
1..1830
   /organism="Arabidopsis thaliana"
   /mol_type="mRNA"
   /cultivar="ecotype H51"
   /db_xref="taxon:3702"
   /chromosomes="4"
   /map="between mi51 and mi122"
1..1830
   /gene="FRI"
1..1830
   /gene="FRI"
   /note="late flowering gene"
   /codon_start=1
   /product="FRIGIDA"
   /protein_id="AG23415.1"
   /db_xref="GI:10801175"
   /translation="MSNYPPTVAAGFTTANPLLRHQHSEQRRELPRKIVETESTMD
ITGQSKQFQFKSIDELAAFSVAVETFKRFDDLRKHIESIENAIIDSKLESNGVILA
ARNNPHQMLSPNNVSVETVTSQSOEIVPETSNTKEGRCMELMCSKGLRY
IVANI SDOAKLMEIIPSAKLAKPAKFLVDCIGKFLQGRRAFTKESPMSSARQVSL
LILESLLMPDRGKGVKIESWKDEASTAAVAKRLMTEGGLAAAEKMDARGLILL
VACPGVPSNFRSTDLLIRMSGNEIAGALKRSQFLVPWVGIVESIRKGMHIEAL
EMWTFGMDKFSALVLITSFLMKSNEFERAKRQSPFLAKSAATKQLAVLSSVMQ
CMETHKLDPAKLPQWIKQEVISLEKDTLQDKEMEEKARSLSLMEEAALAKEMYNQ
QIKRPLSPMEPPVTSYSPIYDRSPFPQDDDDQDEISALVSYLGPSTPPHRS
RRSPETWPLPHGGLGRSVYAYEHLAPNSYSPGHGRLHROYSPSLVHGQRHPLQYSP
PIHGQOLPYGIQVRVHSPEERSPEYVLGLSNQSPRSNSSLDPK"
BASE COUNT   540 a 379 c 436 g 475 t
ORIGIN
Alignment Scores:
Pred. No.:    1..96e-195      Length:    1830
Score:        3104.00        Matches:    609
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:   100.00%      Indels:    0
DB:            8            Gaps:      0

US-09-890-475-1 (1-609) x AF228500 (1-1830)

QY      1 MetSerAsnTyrProProThrValAlaAlaGlnProThrThrThrAlaAsnProLeuLeu 20
Db      1 ATGTCCAAATTATCCACCGAGGTGCGCGGCAACCCCAACCAACGACGGCGAATCCACTGCTG 60

QY      21 GlnArgHisGlnSerGluGlnArgArgGluLeuProLysIleValGluThrGluSer 40
Db      61 CAGCGACATCAATCTCTGAACAGCGAGGAGAGAAATACCGAAGATTGCGAAACAGAGTCT 120

```

```

41 ThrSerMetAspIleThrIleGlyGlnSerLysGlnProGlnPheLeuLysSerIleAsp 60
Db      121 ACARGATATGGACATATACGATCGTCAATCTAAGCAGCCTCAATTTTGAATTCATGAC 180

QY      61 GluLeuAlaAlaPheSerValAlaValGluThrPheLysArgGlnPheAspAspLeuGln 80
Db      181 GAATTAGCTGCGTTTTCAGTTGCAAGTGAACCAATCAACGCCCAATTCGATGATCTTCAG 240

QY      81 LysHisIleGluSerIleGluAsnAlaIleAspSerLysLysLeuGluSerAsnGlyValVal 100
Db      241 AAGCACATCGAGTCGAATCAAAACCAATTAATTCACCAACTCGAGAGTAACGGGTGTC 300

QY      101 LeuAlaAlaArgAsnAsnAsnPheHisGlnProMetLeuSerProProArgAsnAsnVal 120
Db      301 CTCGCGCGCGCAACAATAATTTCCATCAGCGATGTTATCGCTCCGCGGAACAATGTA 360

QY      121 SerValGluThrThrValThrValSerGlnProSerGlnGluIleValProGluThrSer 140
Db      361 TCTGTAGAACCAACCGTCACGTGTGAGCCCAACCGTCTCAGGAGATTGTACCGGAGACGTCG 420

QY      141 AsnLysProGluGlyGlyArgMetCysGluLeuMetCysSerLysGlyLeuArgLysTyr 160
Db      421 AATAAACCGAGGGGGGACGATGTGTGAGTTGATGTGTAGCAAGGCTCGCGTAATATAC 480

QY      161 IleTyrAlaAsnIleSerAspGlnAlaLysLeuMetGluGluIleProSerAlaLeuLys 180
Db      481 ATATACGCGAATATCTCTGATCAAGCTAAGTTAATGGAAGAGATTCCTCAGCTTTGAAA 540

QY      181 LeuAlaLysGluProAlaLysPheValLeuAspCysIleGlyLysPheTyrLeuGlnGly 200
Db      541 TTGGCCNAGGAGCCAGCAAGTTTGTATTGATTTGCAAGTTTACTTACAAAGG 600

QY      201 ArgArgAlaPheThrLysGluSerProMetSerSerAlaArgGlnValSerLeuLeuIle 220
Db      601 CGTAGAGCATTTTACTTAAAGAGTCGCTATGAGCTCTCGAGACAAGTTTCCTCTTATA 660

QY      221 LeuGluSerPheLeuLeuMetProAspArgGlyLysGlyLysValLysIleGluSerTrp 240
Db      661 CTGGAGTCTTTTCTTCTTAATGCTGATCGTGTAAAGGAGGTAAGATTGAGAGTTGG 720

QY      241 IleLysAspGluAlaGluThrAlaAlaValAlaTrpArgLysArgLeuMetThrGluGly 260
Db      721 ATAAAGATGAGCGGAGACGCTGCTGTTGTTGGAGGAAAAGTTGATGAGCTGAAGGA 780

QY      261 GlyLeuAlaAlaGluLysMetAspAlaArgGlyLeuLeuLeuValAlaCysPhe 280
Db      781 GGATTAGCTGCGGCTGAGAAAATGGATGCAAGGGGTTTGCTTTTACTAGTTGCTGTTT 840

QY      281 GlyValProSerAsnPheArgSerThrAspLeuLeuAspLeuIleArgMetSerGlySer 300
Db      841 GGTGTTCTCTTCAAACTTTAGGAGTACAGATTGCTGGAATTGATAAGGATGAGTGGTTCG 900

QY      301 AsnGluIleAlaGlyAlaLeuLysArgSerGlnPheLeuValProMetValSerGlyIle 320
Db      901 AATGAGATTGCGGTCGCTTTTGAACGCTCACAGTTTCTTCTCCCTATGGTCTCAGGTATA 960

QY      321 ValGluSerSerIleLysArgGlyMetHisIleGluAlaLeuGluMetValTyrThrPhe 340
Db      961 GTTGAATCAAGTATCAACGCTGGAAATGCAATATGAAGCTCTTGAGATGGTTTATACCTTT 1020

QY      341 GlyMetGluAspLysPheSerAlaAlaLeuValLeuThrSerPheLeuLysMetSerLys 360
Db      1021 GGCATGAGAGTAAGTTTTCAGCTGCTCTAGTTCTTAACTTCAATCTTAAAGATGAGCAAG 1080

QY      361 GluSerPheGluArgAlaLysArgLysAlaGlnSerProLeuAlaPheLysGluAlaAla 380
Db      1081 GAGTCAATTTGAGAGGGCAAAACGGAAGCCAGTCACCGCTGCGCATTTAAAGAAAGCGGCT 1140

QY      381 ThrLysGlnLeuAlaValLeuSerSerValMetGlnCysMetGluThrHisLysLeuAsp 400
Db      1141 ACAAGCAGCTAGCTGTGTTATCATCATGTTATGAGTGTATGAGTGTATGAGACTCACAAGTAGAT 1200

QY      401 ProAlaLysGluLeuProGlyTrpGlnIleLysGluIleValSerLeuGluLysAsp 420

```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

DM protein - nucleic search, using frame_plus_p2n model

Run on: August 25, 2003, 08:48:34 ; Search time 6375 Seconds
(without alignments)
3908.073 Million cell updates/sec

Title: US-09-890-475-1

Perfect score: 3104

Sequence: 1 MSNPPTVAQPTTANPL.....RYLGLSNQSPRNSSLDPK 609

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n model -DEV=xlh
-Q=/cgn2/1/USPFC_spool/US0890475/runat_14082003_090349_19545/app_query.fasta_1.775
-DB=genEmbl -QMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US0890475@cgn 1 1 3963 @runat_14082003_090349_19545 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl.*

- 1: gb.ba.*
- 2: gb.htg.*
- 3: gb.in.*
- 4: gb.om.*
- 5: gb.ov.*
- 6: gb.pat.*
- 7: gb.ph.*
- 8: gb.pl.*
- 9: gb.pr.*
- 10: gb.ro.*
- 11: gb.sts.*
- 12: gb.sy.*
- 13: gb.un.*
- 14: gb.vi.*
- 15: em.ba.*
- 16: em.fun.*
- 17: em.hum.*
- 18: em.in.*
- 19: em.mu.*
- 20: em.om.*
- 21: em.ov.*
- 22: em.pat.*
- 23: em.ph.*
- 24: em.pl.*
- 25: em.ro.*
- 26: em.sts.*
- 27: em.un.*
- 28: em.un.*

- 29: em.vi.*
- 30: em.htg.hum.*
- 31: em.htg.inv.*
- 32: em.htg.other.*
- 33: em.htg.mus.*
- 34: em.htg.pln.*
- 35: em.htg.rod.*
- 36: em.htg.mam.*
- 37: em.htg.vrt.*
- 38: em.sy.*
- 39: em.htgo.hum.*
- 40: em.htgo.mus.*
- 41: em.htgo.other.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	3104	100.0	1830	8	AF228500 Arabidops
2	3104	100.0	2257	6	AX032763 Sequence
3	2881	96.0	3761	6	AX032762 Sequence
4	2881	96.0	3761	8	AF228499 Arabidops
5	2975	95.8	3150	8	AY198403 Arabidops
6	2975	95.8	3261	8	AY198402 Arabidops
7	2975	95.8	3739	8	AY198401 Arabidops
8	2969	95.7	3301	8	AY198405 Arabidops
9	2950	95.0	3758	8	AY198404 Arabidops
10	2936	94.6	91849	8	AF058919 Arabidops
11	2936	94.6	197975	8	AL61472 Arabidops
12	931	30.0	558	8	AY092681 Arabidops
13	931	30.0	558	8	AY092683 Arabidops
14	931	30.0	558	8	AY092691 Arabidops
15	931	30.0	558	8	AY092693 Arabidops
16	927	29.9	558	8	AY092677 Arabidops
17	927	29.9	558	8	AY092679 Arabidops
18	927	29.9	558	8	AY092694 Arabidops
19	927	29.9	558	8	AY092685 Arabidops
20	927	29.9	558	8	AY092696 Arabidops
21	921	29.7	558	8	AY092678 Arabidops
22	921	29.7	558	8	AY092680 Arabidops
23	921	29.7	558	8	AY092682 Arabidops
24	921	29.7	558	8	AY092684 Arabidops
25	921	29.7	558	8	AY092685 Arabidops
26	921	29.7	558	8	AY092686 Arabidops
27	921	29.7	558	8	AY092687 Arabidops
28	921	29.7	558	8	AY092688 Arabidops
29	921	29.7	558	8	AY092689 Arabidops
30	921	29.7	558	8	AY092690 Arabidops
31	921	29.7	558	8	AY092692 Arabidops
32	845	27.2	511	8	AY176673 Brassica
33	658	21.2	503	8	AY092437 Arabidops
34	658	21.2	503	8	AY092438 Arabidops
35	658	21.2	503	8	AY092439 Arabidops
36	658	21.2	503	8	AY092440 Arabidops
37	658	21.2	503	8	AY092441 Arabidops
38	658	21.2	503	8	AY092442 Arabidops
39	658	21.2	503	8	AY092443 Arabidops
40	658	21.2	503	8	AY092444 Arabidops
41	658	21.2	503	8	AY092445 Arabidops
42	658	21.2	503	8	AY092446 Arabidops
43	658	21.2	503	8	AY092447 Arabidops
44	658	21.2	503	8	AY092448 Arabidops
45	658	21.2	503	8	AY092449 Arabidops

ALIGNMENTS

RESULT 1